



Substance Abuse and Mental Health Services Administration

Center for Substance Abuse Prevention

A Guide for Evaluating Prevention Effectiveness



HV
5825
G84
1998

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Substance Abuse and Mental Health Services Administration
Center for Substance Abuse Prevention

Technical Report



HV
4999.2
84
G
1998

DEC 28 2009

CSAP Technical Report

ADDICTION RESEARCH CENTER
BLDG 10, 10 CENTER DR
BETHESDA, MD 20892-1150
NIH LIBRARY

A Guide for Evaluating Prevention Effectiveness

Center for Substance Abuse Prevention
National Center for the Advancement of Prevention
(301) 984-6500

Contents

Foreword	iii
Introduction	1
Chapter 1: Intervention Types and Their Stages of Development	3
Chapter 2: Frameworks for Evaluation	13
Chapter 3: Outcome Evaluation Designs	19
Chapter 4: Prerequisites to Data Collection	25
Chapter 5: Common Data Collection Methods	31
Chapter 6: Concepts in Data Analysis	41
Chapter 7: Conclusions and Recommendations	49
References	51
Selected Related SAMHSA/CSAP Publications	52
Appendix	53
Glossary	55

List of Figures

Figure 1. Focus of Research During the Course of a Comprehensive Evaluation	14
Figure 2. Evaluation Targets	15
Figure 3. Experimental Design: Simple Pretest-Posttest Control Group	20
Figure 4. Experimental Design: Multiple Group Pretest-Posttest Control Group	20
Figure 5. Experimental Design: Posttest-Only Control Group	21
Figure 6. Experimental Design: Delayed Treatment	21
Figure 7. Quasi-Experimental Design: Nonequivalent Comparison Group	22
Figure 8. Quasi-Experimental Design: Time Series	23
Figure 9. Quasi-Experimental Design: Time Series With Comparison Group	23
Figure 10. Research Methods Employed in Comprehensive Program Evaluation	33

List of Tables

Table 1. Individual-Oriented Interventions To Reduce Substance Abuse	5
Table 2. Policy Interventions To Reduce Substance Abuse Problems	6
Table 3. Evaluation Questions and Data Collection Methods Suited to Different Stages of Development	9
Table 4. Threats to Internal Validity	18
Table 5. Measures for Process and Outcome Evaluations	32
Table 6. Common Data Collection Methods	37
Table 7. Sample Rating Scores on Scale of 1 (Extremely Poor) to 10 (Excellent)	42
Table 8. Contingency Table (Example)	44

Introduction

This Technical Report is intended to be used as a basis for training evaluator-practitioner teams to collaborate in the design and implementation of useful evaluations. It is organized around the idea that a useful evaluation depends on prevention evaluators working closely with practitioners to design feasible evaluations. Practitioners should then support evaluators in their efforts to assess accurately the nature and effects of prevention programs. This document presents both general materials on how evaluation fits into the practice of prevention and technical details of implementing an evaluation plan. The combination will give prevention staff common ground on which to build.

- Chapter 1 of this guide asserts that useful evaluations are tailored to intervention type and stage of development. This information will be useful to preventionists as well as evaluators.
- Chapter 2 describes types of evaluation and evaluation outcomes. It also discusses the problem of credibly attributing observed effects to the prevention intervention. Practitioners and evaluators alike should be familiar with this basic information.
- Chapter 3 outlines several experimental and quasi-experimental evaluation designs. Though of limited value to experienced evaluators, the information will be useful to prevention administrators who may lack understanding of various research designs.

- Chapter 4 discusses critical, practical issues that should be addressed before beginning data collection. Evaluators inexperienced in applied research should review these recommendations thoroughly. Practitioners will find this chapter useful because it details the rights and responsibilities of program personnel and clients in the evaluation process.
- Chapter 5 summarizes basic quantitative and qualitative methods of evaluation. Both practitioners and evaluators should be familiar with these methods.
- Chapter 6 concerns concepts in data analysis with which prevention administrators should be familiar. It notes how some types of data analyses are better suited to specific types of research questions and different stages of program development.
- Chapter 7 draws conclusions and makes recommendations.
- The appendix offers a list of resource books on evaluation. The glossary that follows the appendix will be particularly useful to practitioners unfamiliar with evaluation terms and processes.

Chapter 1

Intervention Types and Their Stages of Development

In order to obtain meaningful information about any prevention intervention, evaluators must select methods that are appropriate for the type of intervention they are evaluating. They must also ask evaluation questions that correspond to the intervention's stage of development. Mismatch between these aspects of the intervention and its evaluation will produce friction and may make both the intervention and the evaluation less useful.

Classifying an intervention according to type and stage of development is the first step in tailoring evaluation methods to the needs of the program. This chapter addresses this critical first step. It begins with a brief discussion of two basic prevention approaches—individual-oriented interventions and policy interventions. It concludes by elaborating on the concept of intervention stage of development, providing sample questions that are appropriate for each stage, and listing examples of research methods that are often used in particular stages.

Approaches to Prevention

Traditionally, the prevention field has emphasized individual-oriented approaches to prevention. After all, smoking tobacco, drinking alcohol, and taking drugs are individual behaviors. It makes sense to intervene with individuals. In fact, interventions with individuals are the most prevalent, and they are relatively easy to implement. Frequently these types of interventions provide services such as employee assistance programs and afterschool alternatives programs. Other interventions attempting to change individual-level behavior include

school-based education programs and “drive safe” media campaigns.

There is growing research evidence, however, that policies can change the environment so as to change individual behavior and to reduce substance-related problems. Policy interventions may be defined as those affecting the social, economic, and regulatory environment around substance use. They attempt to change the environment in such a way that substances become less available and/or more expensive. Policy interventions also change norms, values, and expectations so that they are less supportive of substance use that results in health and social problems. Policy interventions include formal changes in laws at the Federal, State, or local level, as well as changes in institutions (e.g., schools, law enforcement agencies, retail establishments, families).

Interventions that try to reduce the use of alcohol, tobacco, and illicit drugs by focusing entirely on individual knowledge, attitudes, beliefs, and behavior produce less long-term success in decreasing substance-related problems than do policy approaches (Bangert-Drowns, 1988; Clayton, Cattarello, Day, & Walden, 1991; Moskowitz, 1989; Ringwalt, Ennett, & Holt, 1991; Rosenbaum, Flewelling, Bailey, Ringwalt, & Wilkinson, 1994; Rundall & Bruvold, 1988; Tobler, 1992). Individually focused prevention approaches show the most promise when implemented in the context of a larger community change effort (e.g., Holder et al., 1997; Pentz et al., 1989; Perry et al., 1996; Wagenaar, Murray, Wolfson, Forster, & Finnegan, 1994).

Ultimately, policy (or environmental) approaches anticipate that individual decisions to use substances will change. Thus, the distinction between individual and policy approaches is not always clear-cut. For example, sometimes law enforcement strategies to deter substance-related behavior (e.g., sobriety checkpoints) are considered an individual approach because they are intended to persuade individuals not to engage in problem behaviors. On the other hand, these strategies can be seen as one way of establishing an environment in which these problem behaviors are clearly unacceptable.

Table 1 presents examples of interventions aimed at reducing problem behavior through changes at the individual level. Table 2 provides examples of policy strategies for reducing substance use and related problems in the general population.

Individual and policy interventions require somewhat different forms of evaluation, as will be discussed in the chapters that follow. Regardless of intervention type, however, evaluation involves the collection of data in ways that allow evaluators and preventionists to assess and improve the ways in which the intervention is conducted. These methods are fairly well understood, but the specific outcomes measured, study samples selected, and statistical analyses employed are likely to be more similar within intervention type. Evaluations aimed at changing individuals will measure the attitudes, beliefs, and/or behaviors of the individuals exposed to the intervention. In contrast, evaluations of policy interventions will measure community-level and systems-level changes.

Some of the strategies listed in tables 1 and 2 have extensive evidence of effectiveness. Research has not yet established the level of effectiveness for others. It is clear from inspecting table 2, however, that strategies for dealing with illicit drugs are far less available and less well developed than are strategies relevant to alcohol and tobacco. One reason for this disparity is that because alcohol and tobacco are legal substances, there are many more opportunities for reducing consumption and problems through regulatory approaches. Nonetheless, policy approaches to illicit drugs other than conventional law enforcement efforts are emerging, and the knowledge base regarding the effectiveness of these approaches is growing (Bureau of Justice Assistance, 1993; Davis & Lurigio, 1996; Green, 1996; Roehl, Wong, Andrews, Huitt, & Capowich, 1995).

Stages of Development

In addition to recognizing the type of intervention approach, good evaluation methodologies consider stage of development. As prevention efforts mature, the questions posed by the evaluation and the methods used to address them will change. Evaluations during the earlier

Table 1. Individual-Oriented Interventions To Reduce Substance Abuse

<p>School-Based</p> <ul style="list-style-type: none"> • Resistance-skills training program designed to increase youths' ability to withstand the pressure or temptation to use alcohol, tobacco, or drugs • Skills-building program increasing social and academic abilities • Alcohol, tobacco, and drug educational program teaching students about the dangers and risks associated with use and fostering a more accurate perception of norms 	<p>Church-Based</p> <ul style="list-style-type: none"> • Group discussion or support sessions centered around personal difficulty management (e.g., stress, grief, divorce) • Series of sermons on the dangers of smoking • Testimony during church services from members giving up the use of substances
<p>Work-Based</p> <ul style="list-style-type: none"> • Employee assistance (EA) program instructing employees how to handle stress at work and at home • EA program teaching employees about the risks and dangers associated with abuse • EA program enabling parents to prepare their children for self-care through discussions about self-care issues (safety, decisionmaking, communication, substance abuse) 	<p>Community-Based</p> <ul style="list-style-type: none"> • Afterschool program for latchkey children • Parent-to-parent network helping all parents in a community to supervise children • Mentoring program exposing youths to positive adult role models and encouraging high academic and professional aspirations
<p>Family-Based</p> <ul style="list-style-type: none"> • Skills-building and educational program for parents and youth to improve communication and increase knowledge about substance-related issues • Parenting skills-building program for single and/or young parents who may need assistance dealing with needy or troubled children • Family therapy to improve communication and attachment in families of delinquent youths 	<p>Health Care-Based</p> <ul style="list-style-type: none"> • Training of health care providers to detect the signs and symptoms of substance abuse • Discussion of the dangers of substance abuse in every health service encounter
<p>College-Based</p> <ul style="list-style-type: none"> • Discussion groups examining advertisement and promotional strategies of the alcohol and tobacco industries • Support groups for adult children of alcoholics to examine issues related to parental alcoholism and stressful situations both on campus and at home • Discussion of the dangers of substance abuse in every nonacademic counseling encounter 	<p>Other</p> <ul style="list-style-type: none"> • Driver education classes that include attention to the risks associated with alcohol- and drug-impaired driving and penalties for driving under the influence • Court-ordered educational or treatment programs for impaired drivers • Media-sponsored "drive safe" campaigns • Elderly-outreach program to assist with stressful life circumstances (e.g., retirement, grief, living alone)

Table 2. Policy Interventions To Reduce Substance Abuse Problems

Alcohol	
<p><i>Public Policies</i></p> <ul style="list-style-type: none"> • Excise taxes • Limits on hours or days of sale • Restrictions on density, location, and type of outlets • Mandatory server training and licensing • Dram shop and social host liability • Restrictions on advertising and promotion • Mandatory warning signs and labels • Restrictions on consumption in public places • Restrictions on happy-hour sales • Prevention of preemption of local control of alcohol regulation • Minimum drinking age • Keg registration ordinances • Enhancement of drivers' licenses (to indicate age clearly and prevent fraud) • Ban on home deliveries • Compulsory compliance checks for minimum purchase age and administrative penalties for violations • Establishment of minimum age for sellers • 0.00 blood alcohol content (BAC) for young drivers • 0.08 BAC for adult drivers • Administrative license revocation for impaired drivers 	<p><i>Organizational Policies</i></p> <ul style="list-style-type: none"> • Warning posters (<i>businesses</i>) • Restrictions on alcohol advertisements (<i>media</i>) • Restrictions on alcohol use at work and work events (<i>businesses</i>) • Restrictions on sponsorship of special events (<i>communities, stadiums</i>) • Police walkthroughs at alcohol outlets • Undercover outlet compliance checks (<i>law enforcement agencies</i>) • Responsible beverage service policies (<i>outlets</i>) • Mandatory checks of age identification (<i>businesses</i>) • Server training (<i>businesses</i>) • Incentives for checking age identification (<i>businesses</i>) • Restrictions on sales to those accompanied by individuals under age 21 (<i>businesses</i>) • Prohibition of alcohol on school grounds or at school events (<i>schools</i>) • Enforcement of school policies (<i>schools</i>) • Prohibition of beer kegs on campus (<i>colleges</i>) • Establishment of alcohol-free dormitories and campuses (<i>colleges</i>) • Establishment of enforcement priorities against adults who illegally provide alcohol to youth • Sobriety checkpoints (<i>law enforcement agencies</i>) • Media campaigns about enforcement efforts (<i>media</i>) • Safe ride programs (<i>businesses</i>) • Identification of source of alcohol consumed prior to driving-while-intoxicated arrests (<i>law enforcement agencies</i>)

Table 2. Policy Interventions To Reduce Substance Abuse Problems (continued)

Tobacco	
<i>Public Policies</i> <ul style="list-style-type: none"> • Excise taxes • Tobacco sales licensing system • Prohibition of smoking in public places • Prevention of preemption of local control of tobacco sales • Restrictions on advertising and promotion • Ban on vending machines • Compulsory compliance checks for minimum purchase age and administrative penalties for violations • Minimum age of sale of 18 • Warning labels • Mandatory seller training • Ban on self-service sales (all tobacco behind the counter) • Minimum age for sellers 	<i>Organizational Policies</i> <ul style="list-style-type: none"> • Establishment of smoke-free settings (<i>restaurants, workplaces, hospitals, stadiums, malls, day care facilities</i>) • Counteradvertising (<i>media</i>) • Restrictions on sponsorship of special events (<i>communities, colleges, stadiums</i>) • Prohibition of tobacco use on school grounds, in buses, and at school events (<i>schools</i>) • Enforcement of school policies (<i>schools</i>) • Mandatory checks of age identification (<i>businesses</i>) • Seller training (<i>businesses</i>) • Incentives for checking age identification (<i>businesses</i>) • Undercover shopper or monitoring program (<i>businesses</i>)
Other Drugs	
<i>Public Policies</i> <ul style="list-style-type: none"> • Control of production and distribution • Zoning and building codes that discourage drug activity and penalties for property owners who fail to address known drug activity • Mandated school policies 	<i>Organizational Policies</i> <ul style="list-style-type: none"> • Employer policies (<i>businesses</i>) • Surveillance of high-risk public areas (<i>law enforcement agencies, neighborhood watch groups</i>) • Enforcement of zoning and building codes (<i>law enforcement agencies, building authorities</i>) • Appropriate design and maintenance of parks, streets, and other public places (e.g., lighting, traffic flow) (<i>city agencies, housing authorities</i>) • Enforcement of school drug policies (<i>schools</i>)

Note: Institutions that can develop suggested institutional policies are indicated in parentheses.

phases of an intervention tend to focus research questions around the nature of its implementation. Such evaluations are called *process evaluations*. Evaluations during later phases center research questions around finding evidence of the intervention's effect on desired substance abuse-related outcomes. Such evaluations are called *outcome evaluations*.

Regardless of the intervention's stage of development, some evaluation questions should anticipate issues beyond the immediate concern. For example, even though a new prevention effort may be in its infancy and most evaluation questions center around the nature of implementation and/or service delivery, some research questions should focus on anticipating and detecting initial or preliminary outcomes. Similarly, a prevention effort in later stages of development (with research questions focused around outcomes) should continue to collect data about the nature of implementation and/or service delivery in order to detect process issues that may contribute to unanticipated outcomes.

Of course, the stage of development notion does not completely capture the dynamic nature of intervention development. In reality, prevention efforts are highly dependent on personnel, funding, and other organizational elements, and development is not always sequential. Prevention personnel may find themselves in several stages simultaneously. A given step may be repeated until success is achieved, or it may be skipped altogether. An awareness of this general developmental course, however, can be useful in tailoring the evaluation appropriately to the stage of development.

Differentiating prevention efforts according to their stage of development should (1) make expectations for the short-term accomplishments of the intervention more reasonable; (2) provide information that can help prevention efforts move further along in their development; (3) reduce evaluation costs by limiting evaluation activities to those needed to answer the questions of greatest relevance; and (4) increase the usefulness of evaluation activities by focusing on pertinent questions.

Table 3 lists the stages of intervention development, provides examples of the types of eval-

uation questions likely to be asked at each stage, and suggests different evaluation methods that might be employed to answer those questions. This classification combines ideas from several sources (Florin, Mitchell, and Stevenson, 1993; French & Bell, 1984; Goodman & Wandersman, 1994; Gottfredson, 1984; Kibel, 1994; National Institute on Drug Abuse, 1988, 1991; Sechrest & Figueredo, 1993).

The wide array of qualitative and quantitative evaluation methods shifts from description and exploratory analyses in the earlier stages of the evaluation to more comparative and confirmatory analyses in later stages. Thus, a mix of evaluation methods is required and, ultimately, most useful. Consequently, practitioners and evaluators should:

- Select evaluation questions appropriate for the intervention's stage of development;
- Select an appropriate mix of data collection techniques; and
- Assess the nature of implementation, as well as its effects on substance abuse-related outcomes.

Summary

This chapter suggests that appropriate evaluation methodologies stem from a recognition of an initiative's characteristics. Evaluations must be suited to the type of initiative as well as to its stage of development. Frequently, evaluations will be most useful if they evolve in tandem with the prevention effort as it moves through various stages of development. Process evaluation questions are most common during earlier stages of an initiative. Outcome evaluation questions are more appropriately asked of a mature program. However, the groundwork for a rigorous outcome evaluation must be laid earlier in the life of a project. It is therefore necessary to anticipate and plan for outcome evaluation well before the program is ready for this level of scrutiny. The immediate information needs of a developing project, however, will center around issues other than outcomes during earlier stages of development. Similarly, process information about program implementation should be collected throughout the life of the program.

Table 3. Evaluation Questions and Data Collection Methods Suited to Different Stages of Development

Stage of Development/Tasks	Sample Evaluation Questions	Examples of Appropriate Data Collection Methods
<p><i>Initiation Stage</i></p> <ul style="list-style-type: none"> – Mobilize support and resources – Develop intervention capacity – Begin documenting intervention's development – Begin thinking about evaluating the intervention 	<p>What is the level of interest or concern about the problem in the community?</p> <p>Who is involved in developing the intervention?</p> <p>What resources (human and other) are available and needed?</p> <p>Is there sufficient political support for this effort?</p>	<p>Community survey to assess concerns</p> <p>Content analysis of newspaper and other media accounts</p> <p>Survey of participants' perceptions of "organizational climate" of the group and disposition for implementing the initiative</p> <p>Observation of the community mobilization process</p> <p>Organizational analysis to assess organizations and persons participating in awareness and mobilization efforts, noting breadth of community sectors represented and number of key community leaders involved</p>
<p><i>Planning Stage</i></p> <ul style="list-style-type: none"> – Determine nature and extent of problem – Investigate alternative approaches – Develop theoretical framework – Define intervention goals – Develop initial implementation plan – Develop preliminary evaluation plan – Begin evaluation 	<p>What are the characteristics and needs of the target population?</p> <p>What are the perceived causes of the problem?</p> <p>What is the cultural context in which the intervention will operate?</p> <p>Who is involved in developing the intervention?</p> <p>What are the existing policies that address the problem?</p> <p>Are there specific constituencies that might be inclined to support policy changes?</p> <p>What prevention interventions are available and appear to be acceptable to the community?</p> <p>What obstacles need to be overcome for successful implementation?</p> <p>What are the expectations for the intervention?</p>	<p>Survey and review of archival data on the nature and extent of problems</p> <p>Survey of attitudes and beliefs about the targeted problem and the proposed approaches</p> <p>Analysis of similar previous or existing prevention efforts in the community</p> <p>Counts of organizations and individuals involved in the planning process</p> <p>Content analysis of initial plan, including attention to: (1) the presence, clarity, and reasonableness of theory of action; (2) the match between presumed causes of the problem to be addressed and the approaches to reducing problems; (3) the level and clarity of expectations for the activities; (4) anticipated obstacles and plans to overcome them; and (5) numbers and types of planned activities</p>

Table 3. Evaluation Questions and Data Collection Methods Suited to Different Stages of Development (continued)

Stage of Development/Tasks	Sample Evaluation Questions	Examples of Appropriate Data Collection Methods
<p><i>Pilot-Testing Stage</i></p> <ul style="list-style-type: none"> – Monitor intervention to detect problems – Solve implementation problems – Refine implementation plan – Refine evaluation plan – Continue education 	<p>What are participants'/clients' current behavior, attitude, or understanding about the problem?</p> <p>What resources (anticipated and unanticipated) are used?</p> <p>What obstacles are encountered?</p> <p>Are organizational processes and systems (e.g., communication, decisionmaking, lines of authority) adequate?</p> <p>Is the intervention implemented as intended?</p> <p>What are clients'/participants' reactions to the intervention?</p> <p>What costs are associated with the intervention?</p>	<p>Survey to assess behavior, attitudes, knowledge, or actions (or similar preintervention assessment)</p> <p>Analysis of records on recruitment efforts, attendance patterns, and expenditures</p> <p>Observations of intervention context and/or service delivery</p> <p>Focus groups with policy implementers, service deliverers, and/or service recipients to assess their opinions of the intervention</p> <p>Examination of implementation logs</p> <p>Followup interviews with program dropouts to ascertain causes of attrition</p> <p>Client satisfaction questionnaire</p>
<p><i>Implementation Stage*</i></p> <ul style="list-style-type: none"> – Monitor intervention to detect problems and immediate needs – Monitor support for and satisfaction with intervention – Monitor intervention effects – Monitor capacity – Refine intervention (if necessary) 	<p>Is the intervention implemented as intended?</p> <p>What resources (anticipated and unanticipated) are used?</p> <p>What obstacles are encountered?</p> <p>Are organizational processes and systems adequate?</p> <p>What costs are associated with the intervention?</p> <p>Are staff adequately trained to deliver services or implement policy changes?</p> <p>What changes in behavior, attitudes, knowledge, or actions are evident?</p> <p>What unanticipated effects occur?</p> <p>What are clients'/participants' reactions to the intervention?</p> <p>What is the overall impact on the community and/or target population?</p>	<p>Observation of intervention context and/or service delivery</p> <p>Analysis of program records on recruitment efforts, attendance patterns, and expenditures</p> <p>Examination of intervention implementation logs</p> <p>Focus groups with service providers or policy implementers to assess their opinions of the intervention</p> <p>Followup interviews with program dropouts to ascertain causes of attrition</p> <p>Focus groups to assess changes in attitudes and knowledge as well as opinions</p> <p>Survey to assess changes in behavior, attitudes, knowledge, or actions (or similar postintervention assessment)</p> <p>Client satisfaction questionnaire</p> <p>Analysis of archival records concerning incidence of focal problem</p> <p>Community survey of reactions to the intervention</p>

Table 3. Evaluation Questions and Data Collection Methods Suited to Different Stages of Development (continued)

Stage of Development/Tasks	Sample Evaluation Questions	Examples of Appropriate Data Collection Methods
<p><i>Stabilization Stage*</i></p> <ul style="list-style-type: none"> – Demonstrate value to sponsors, clients, and community – Enhance capacity – Defend budgets and resources in the face of competing claims for resources – Assess and sustain intervention effects 	<p>Do organizational processes and systems remain adequate?</p> <p>Is service delivery or policy implementation maintained and/or improved?</p> <p>Are new costs associated with the intervention?</p> <p>Do staff remain qualified to deliver services or implement policy appropriately?</p> <p>Are new changes in behavior, attitudes, knowledge, or actions observed?</p> <p>Are there any other unanticipated effects?</p> <p>Are positive effects maintained over time?</p> <p>Are benefits worth the cost?</p>	<p>Focus groups to assess changes in attitudes and knowledge as well as opinions of service delivery</p> <p>Focus groups with service providers or policy implementers to assess their opinions of the intervention</p> <p>Community survey to assess reactions to the intervention</p> <p>Analysis of intervention expenditures</p> <p>Assessment of generalizability of findings</p> <p>Analysis of questionnaires, interviews, observations, archival records</p>
<p><i>Dissemination Stage</i></p> <ul style="list-style-type: none"> – Assess projected needs – If considering expansion, <ul style="list-style-type: none"> • expand resources • engage in Initiation Stage tasks 	<p>Are there alternative approaches or additional policies that should be explored and possibly implemented?</p> <p>Should interventions be targeted to new areas?</p> <p>Is expansion worth the cost?</p>	<p>Examination of documentation of decisions to expand the prevention effort</p> <p>Assessments of needs to services provided</p> <p>Cost-benefit analysis of monetary expenditures to documented positive effects</p>

* The tasks and questions associated with the Implementation and Stabilization Stages are not exclusive to the stage and, once posed, tend to be ongoing throughout the course of the program.

Chapter 2

Frameworks for Evaluation

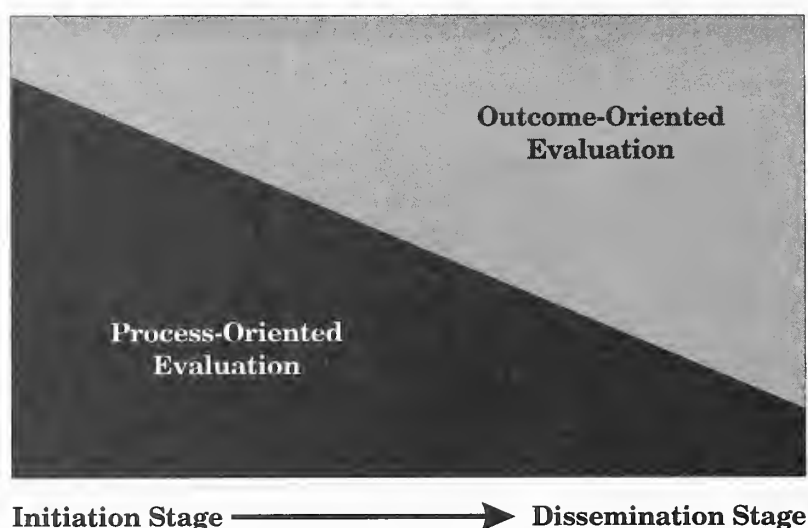
Evaluations should provide information that enables managers and policymakers to make timely decisions concerning appropriate prevention efforts. This chapter discusses the conceptual building blocks necessary for conducting useful evaluations. It begins with a brief discussion of the distinctions between process and outcome evaluations and between intermediate and long-term outcomes, then explains the concept of attribution of effect.

Process Versus Outcome Evaluations

Although there are multiple approaches to evaluation, comprehensive approaches involve the systematic gathering of information about the prevention intervention's operation, as well as its effects. Thus, evaluators conduct both process and outcome evaluations to get a complete understanding of an intervention. Figure 1 illustrates how the proportion of process- to outcome-oriented data collection activities tends to shift in an ongoing evaluation. Note that process evaluation tends to dominate in early phases but continues throughout the course of the evaluation. The emphasis on outcome evaluation tends to increase as the intervention matures.

Process Evaluation

Process evaluation is ongoing assessment and documentation of the planning, development, and implementation phases of an intervention. Largely descriptive, process evaluation can focus on numbers and characteristics of clients, services, and program staff. Process evaluation can also allow implementers to compare inter-



Stages of Program Development

Figure 1.—Focus of research during the course of a comprehensive evaluation.

vention progress with intervention objectives and expectations. If the goals and expectations are not being met, the intervention can be adjusted.

Perhaps the most useful aspect of process evaluation is that it enables evaluators to interpret or understand better the intervention's outcomes. For instance, an ordinance may be passed mandating that all alcohol servers receive training on how to avoid serving alcohol to intoxicated or underage customers. If later outcome evaluation determines that intoxicated and underage patrons were still able to buy alcohol, process evaluation might indicate why the intervention did not have the intended effect. Perhaps not enough servers actually received the training. Perhaps turnover among staff at retail outlets is so great that the training should be repeated more frequently. Perhaps management of retail establishments did not support service staff in their efforts to avoid inappropriate sales. The data collected through process evaluation can provide important information for explaining the causes of high or low performance levels (i.e., outcomes). Methods for collecting data to evaluate the process of a prevention intervention are discussed in chapter 5.

Outcome Evaluation

Often when prevention administrators state that an intervention has been successful, they cite data derived from process evaluation as evidence of this success. That is, they cite observational data on practices, participants' reactions to or evaluations of services, and other types of feedback. While such information is useful, it is important to move beyond this type of assessment. Prevention success ultimately must be defined by outcomes.

In essence, outcome evaluations attempt to find out whether a prevention effort made a difference in the lives of clients or in the community. Specifically, outcome evaluation documents whether a prevention intervention produced the desired effects: for example, changes in behavior or in the number of problem events. These evaluations also look for unexpected outcomes or side effects. Sometimes, outcome evaluations even ask how the benefits of the intervention compared with the costs.

Most policy interventions and many individual-oriented interventions have intermediate and long-term goals. In the long term, these interventions are intended to reduce substance-related health and social problems either in

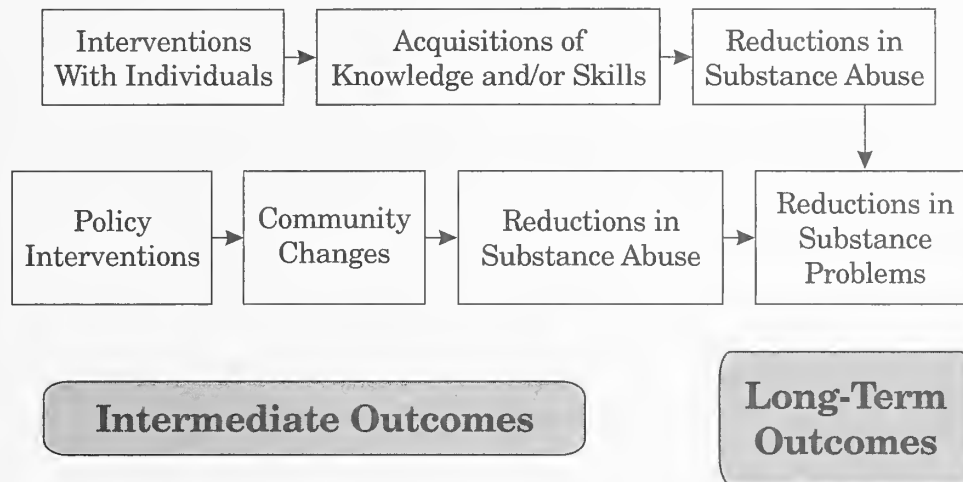


Figure 2.—Evaluation targets.

individuals or in a community. These long-term outcomes are achieved, however, through the accomplishment of intermediate outcomes. Figure 2 illustrates the stages by which prevention interventions alter behaviors and problems at both the individual and community levels.

Examples of intermediate outcome evaluation questions include:

- Did the intervention change attitudes among the targeted population? For example, if the intervention was designed to increase community support for a specific policy or program, did public support actually improve?
- Did the intervention increase skills among the targeted population? For example, if the intervention was designed to increase drug resistance skills, did they actually increase?
- Was the environment actually changed by the intervention? For example, if the intervention was designed to reduce youth access to a substance, did access actually decrease?
- Did the intervention change specific targeted behaviors? For example, if parents were to discuss substance abuse issues with their children more frequently, did they actually do so?
- Did rates of use of the targeted substance decline?

Long-term outcome evaluation questions may include:

- Did use among the targeted population decline?
- Did rates of targeted problems decrease?
 - Was the frequency of the problem lower after the intervention?
 - Did problems decrease more in communities experiencing the intervention than in comparison communities?
 - Did problems decrease more in certain populations than in others?
 - Were reductions in problems sustained, or did beneficial effects dissipate over time?

The answers to these and similar questions can help improve existing interventions, provide the motivation to sustain successful efforts, and prompt other communities to implement similar strategies.

The selection of appropriate intermediate variables is important and requires a thorough understanding of the causation of substance-related problems and the ways in which strategies work to reduce problems. For instance, a family-oriented prevention program may attempt to prevent later substance abuse among youth by improving parenting skills of parents of small children. The intermediate outcome is changes in family management; the long-term outcome is reduced substance abuse by children.

In another example, reduction of the legal blood alcohol level for driving is one policy approach to reducing alcohol-related traffic crashes. An apparently logical intermediate outcome would be an increase in the number of arrests made at the lower blood alcohol level. In fact, few such arrests are likely to be found. Even after the law is changed, police tend to arrest drivers with very high blood alcohol levels because they are easier to detect. Nonetheless, the lowered level does have the effect of reducing alcohol-related crashes—for example, by making people feel that there is a greater threat of arrest when driving with any amount of alcohol. Consequently, people are less inclined to drink and drive. Thus, it is important when designing any evaluation to think through carefully the chain of events that the intervention is likely to cause. Understanding this “causal model” will help ensure that the appropriate outcomes will be considered at each step of the evaluation.

If the intermediate outcome is not in fact causally related to the long-term outcome, the results of the evaluation will not actually predict prevention success. For example, many prevention programs have used increased self-esteem as an outcome measure, even though there is no correlation between self-esteem and substance abuse (Schroeder, Laflin, & Weis, 1993). Evaluators and program planners should carefully examine their assumptions about the relationship between the intermediate and long-term outcomes and, whenever possible, include actual measures of substance use and problems as part of the evaluation. This usually means a reduction in a specific problem or decreased risk of a problem. For example, if the target outcome is self-reported drug use by young people, then such drug use should go down. In some cases, the substance-related outcome itself can be measured. For example, changes in the rate of alcohol-related traffic crashes can occur fairly quickly following policy interventions, making it possible to determine the effectiveness of policy interventions with this focus.

Evaluating both intermediate and long-term outcomes may at first appear to be excessive. After all, we want to reduce substance-related problems, so why not just measure rates of

these problems? Obviously the ultimate goal is to assess the effect of an intervention on health and social problems, but measuring the effects of strategies on problems may be difficult. In some cases, it may take many years before benefits emerge. For example, the life-saving benefits of preventing the sale of tobacco to minors will not emerge for decades. In some cases, the shifts in State and community norms and values prompted by policy strategies will take time to develop. By measuring intermediate outcomes, the relative effectiveness of interventions can be forecasted and determined more quickly. Also, if an intervention is not successful, evaluation of intermediate outcomes helps to identify where it can be strengthened.

Attribution of Effect

The purpose of any evaluation is to determine whether an intervention was effective according to stated goals and as measured by valid outcome criteria. Effectiveness can be measured by (1) comparing postintervention conditions (outcomes) with what would have happened had the intervention not taken place, or (2) comparing one set of postintervention conditions with those of another intervention.

The problem in evaluating an intervention in its own right is that it is not possible to observe what would have happened to a group of people without the intervention or if an alternative intervention had been implemented. Analysts must identify and develop methods of predicting what would have occurred if a specific intervention had not taken place. The population that received the intervention may be compared with a similar population or group that did not receive the intervention, which acts as the control. For example, a community partnership targets four public housing sites, situated in one ward of a large city, for intensive delivery of its youth- and family-focused prevention programs. Public housing sites that are similar in terms of living conditions and resident populations to those targeted for services could serve as comparison sites. By comparing conditions between the two sites before and after the intervention, the amount of change that interventions produce in substance abuse indicators or other desired outcomes can be

measured. Thus, systematically varying interventions across individuals, sites, or time can provide information about the strength of relationships between interventions and outcomes.

Ideally, the intervention is designed as an “experiment.” True experiments require that the units of analysis (for example, community residents or chosen communities) be randomly selected to participate in the intervention. Random assignment occurs when each member of a pool of potential participants has an equal chance of being assigned to the intervention group and the control group. The advantage of a true experimental design is that random assignment ensures that the intervention and control groups do not differ from each other in any systematic way. If differences are found between the two groups using statistical methods, they are most likely due to the intervention.

When a prevention effort is directed at individuals, it is often possible to assign participants randomly in a study. Some evaluations randomly assign classrooms, schools, neighborhoods, or even communities to experimental or control groups. Often, however, random assignment may be too expensive, impractical, or even inappropriate.

Because true experiments often are not possible or are too expensive, prevention evaluations must sometimes rely on what is known as “quasi-experimental design” to isolate the relationship between interventions and outcomes. In the absence of random assignment, the ability to analyze linkages between interventions and outcomes depends on the ability to isolate their interrelationships. For instance, adolescents whose parents are extremely supportive of high academic achievement might be more likely than others to comply with school-based programs of all sorts, including substance abuse prevention, and might also be less at risk for substance use. In this and other instances of apparent intervention-to-outcome linkages there may be no real relationship between participation in program interventions and outcomes. Individuals who participate more in school programs might be found to have more favorable outcomes, such as lower indices of substance use risk. That may not mean that the

intervention caused the outcome. The relationship could be due to unmeasured factors that influence both the participation in interventions and the attainment of outcomes.

Considering the role of factors other than the intervention is also important in evaluating interventions conducted at the community or State level. For example, the number of alcohol-related car crashes may decline following a community prevention intervention. The decline in crashes may, however, be due to an increase in gasoline prices, a new State law increasing penalties for driving under the influence, or a recession that reduces personal incomes. Indeed, the change in auto crashes may be part of a downward trend that preceded the prevention program. The challenge of outcome evaluation is to sort out the effects of other factors so that any changes in behavior can be accurately attributed to the prevention effort itself.

Researchers use the term “internal validity” to indicate that the intervention being tested has an effect independent of other factors that might result in a similar outcome (Campbell & Stanley, 1963). Table 4 highlights some of the problems encountered when researchers try to isolate the effects of an intervention from extraneous factors that may affect the desired outcome. These problems include history, maturation, measurement effects, statistical regression, and selection. It is generally not possible to eliminate all of these problems, but attempts are made to minimize and measure their impact.

Summary

This chapter briefly outlined the differences between process and outcome evaluations and between intermediate and long-term outcomes. It then discussed the problem of accurately attributing effects to prevention interventions. A credible evaluation plan will acknowledge the concepts presented in this chapter by indicating how common evaluation problems will be addressed.

Table 4. Threats to Internal Validity

Threat to Internal Validity	Definition	Example
History	Events or conditions other than the intervention may have an impact on the experimental group and affect outcome measures independent of the intervention.	If a designated driver program is implemented, other factors, such as a change in the price of gasoline or alcohol, or enforcement of stricter legal penalties for driving while intoxicated (DWI), could reduce DWI and related accidents, independent of the designated driver program.
Maturation Effects	Changes in the experimental group unrelated to the intervention may influence outcome measures.	Children tend to use more substances as they get older, independent of external circumstances.
Measurement Effects	The process of collecting information may influence responses on the measures being used to assess outcomes. While completing a survey or test, the respondent's knowledge or attitudes may be altered, and subsequent responses may change solely as a result of having been surveyed, observed, or interviewed. Problems may also occur with archival data because of changes in reporting rather than actual changes in incidence.	Questions about attitudes toward alcohol or drugs on a preintervention survey may make respondents reflect on their alcohol and drug attitudes and change their responses to more socially acceptable ones on a postintervention survey.
Regression to the Mean	When a value is extreme or out of the ordinary, subsequent values tend to be closer to or regress to the mean.	If there is a rash of substance abuse-related incidents in a year, there are likely to be fewer substance abuse-related incidents in subsequent years, with or without the intervention.
Selection Effects	An apparent experimental effect occurs because of an inequality between the treatment and control groups. Differences between the experimental and control groups after an intervention may result from unobserved initial differences between the groups, rather than as a result of the intervention.	A school-based substance abuse prevention program was found to be very effective in that participants reported less drug use than nonparticipants. Students who already had negative attitudes toward drug use, however, were more likely to participate in the program.

Chapter 3

Outcome Evaluation Designs

In this chapter two types of research designs for outcome evaluations are discussed: experimental and quasi-experimental. *Neither type of research design is better or more scientific than the other design type.* Whether a research design can produce credible results depends on how well suited the research design is to the intervention and how the research design is actually implemented within the context of the evaluation. The books listed in the appendix contain lengthier discussions of issues relating to evaluation design.

Experimental Designs

Experimental designs always include random assignment to experimental or control conditions to assist with attribution of effect. Different experimental designs are useful in different situations. Several alternatives are described below.

Pretest-Posttest Control Group Design

In a pretest-posttest control group design, participants are randomly assigned to either a treatment group or a control group. As can be seen in figure 3, the two groups are twice assessed at the same time. Only one group, however, is exposed to an intervention in the period between the two assessments.

By way of example, Quinn County Prevention Coalition operates two innovative programs designed to encourage a healthy lifestyle for youths and to discourage drug use, and it wants to find out which program has a greater effect on youths' behavior. Using a pretest-

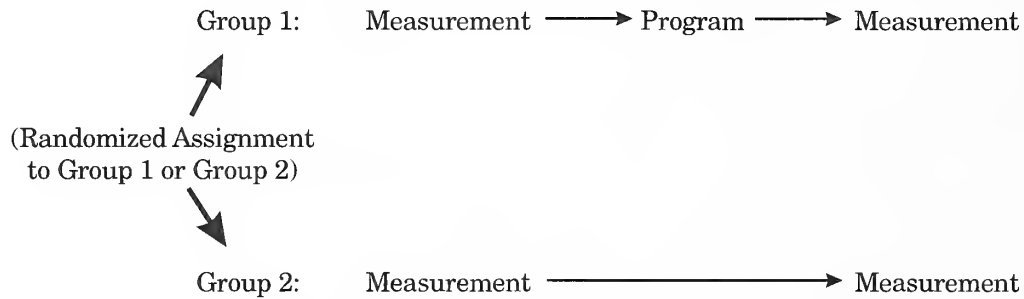


Figure 3.—Experimental design: simple pretest-posttest control group.

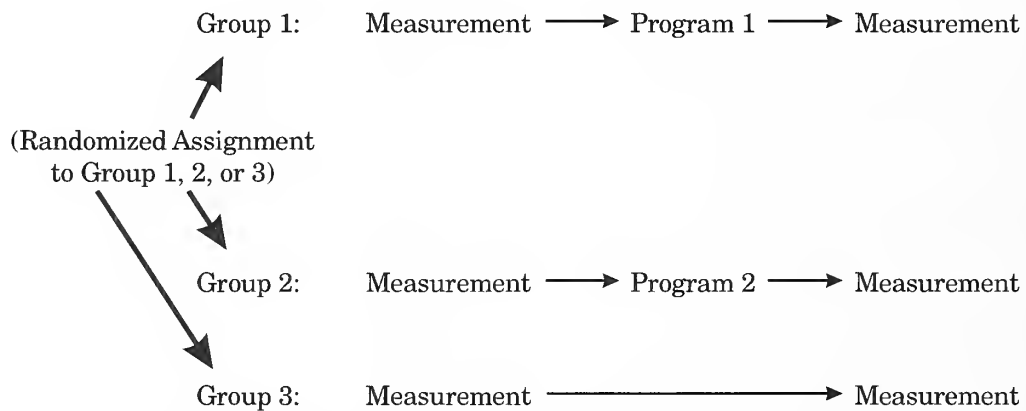


Figure 4.—Experimental design: multiple group pretest-posttest control group.

posttest control group design, sophomores at Washington High School are randomly assigned to three health classes, one for each program and one as a control group.

Before beginning the programs, all students are administered a pretest that measures lifestyle practices and drug use. In one class, the instructor teaches the innovative health prevention curriculum (intervention Group 1); in a second class, 10 “superstars” make 1-hour presentations (intervention Group 2); and in a third class, the teacher conducts the traditional health class (Control Group). At the end of the 5-week period, all students are administered a posttest that measures the attitudes and behaviors targeted for change. Test scores are then compared across classes to determine effectiveness.

Figure 4 presents the Washington High School example. This design enables the evaluators to conclude that whatever differences are measured among the groups, they are most likely due to differences among the programs the students received. Ideally, of course, a longer term followup would be desirable to determine if changes persist over time.

Posttest-Only Control Group Design

In this type of experimental design, participants are randomly assigned to either the intervention group or the control group. The intervention group is given both a pretest and a posttest, and the control group is given only a posttest. Some evaluators believe that this type of experimental design controls for any confusing results that may be produced by a control group pretest. For example, in the absence of exposure

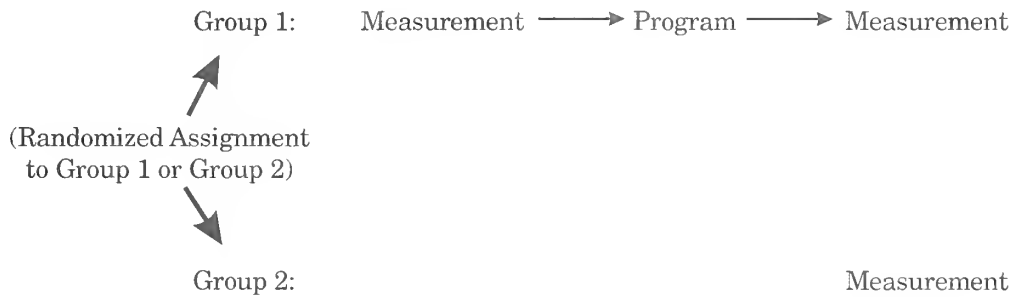


Figure 5.—Experimental design: posttest-only control group.

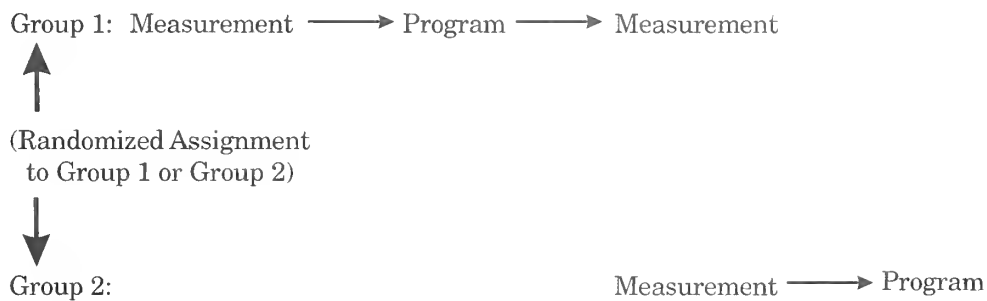


Figure 6.—Experimental design: delayed treatment.

to an intervention, control group test-takers may nevertheless score quite differently on a posttest simply as a result of having taken the test the first time.

Figure 5 depicts this type of true experimental design. In this case, the evaluators assume that the experimental and control groups were similar before the program, but they do not have test scores to prove that assumption.

Delayed Treatment Design

When there is good reason to believe that a program is beneficial and has no harmful side effects, it is difficult to deny services to participants assigned to the control group. The delayed treatment design is planned so that participants are randomly assigned to receive the prevention program services at different times. For example, the new 5-week health curriculum at Washington High School begins on

October 1 in Class A and starts 5 weeks later in Class B. At the end of the first week in November, measurements for Class A and Class B can be compared to assess the effects of the curriculum because it has not yet started in the second class. However, once the program starts in the second class, Class B can no longer serve as a control group for Class A. Figure 6 portrays an example of this type of true experimental design.

Quasi-Experimental Designs

In many situations, an experimental design is inappropriate, too costly, or impractical—for example, when an intervention is applied to a whole community. While there is a wide range of quasi-experimental designs, most are variations of two basic approaches. The first approach involves the use of a comparison group that was not exposed to the intervention. The second approach uses the experimental

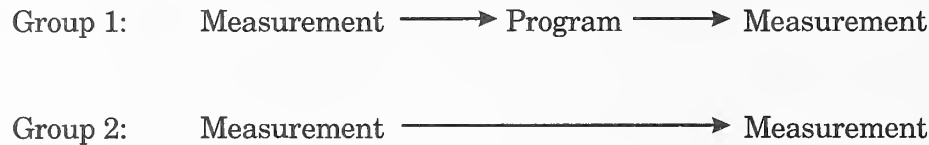


Figure 7.—Quasi-experimental design: nonequivalent comparison group.

group as its own control. Thus, quasi-experimental designs do not rely on random assignment but employ other research techniques, as well as statistical analyses, to assist with attribution of effect. Three types of quasi-experimental designs, discussed below, are commonly used.

Nonequivalent Comparison Group

This type of design compares the outcome of the group that receives the program with that of a comparison group that does not receive the program. A comparison group is selected because it appears to be similar in many respects to the group receiving the program, but assignment to the experimental group versus the comparison group is not random. Statistical analysis is used to control for any preexisting differences between the two groups. For example, a community might wish to test the effectiveness of a mandatory responsible beverage service program for bars and restaurants as a means of reducing impaired driving. Following the implementation of the program, evaluators could compare the occurrence of impaired driving crashes in the community with that of a similar community that did not have such a program.

In this case, it is important to select a community as similar as possible to the experimental community based on certain preselected criteria (e.g., race, socioeconomic status), although a pretest allows evaluators to control statistically for some initial differences between groups. If the communities differ in the initial drinking and driving rates, statistical adjustments for these preintervention differences can be made to ensure an accurate comparison of the communities at the posttest. In order for the pretest to be used in this manner, it must contain measures of all relevant dimensions along

which the intervention and comparison groups differ.

A weakness of this quasi-experimental design is that it is difficult to be certain that all relevant characteristics have been measured well enough to control statistically for all plausible causes of posttest differences other than the program itself. In general, the lack of randomized assignment in this type of design allows for greater threats to causal inferences. Figure 7 presents the nonequivalent comparison group design.

Time-Series Design

This design requires multiple measures of the desired outcome in years or months before and after the intervention. Measurements taken before and after are then compared and analyzed for trends and changes in trends. This type of research design allows for lags in the effects of the intervention. Archival data such as impaired driving crash and arrest rates are particularly useful for this type of analysis because they are maintained for long periods of time. Figure 8 depicts this design.

A time-series design would not establish causality if an unplanned event coincided with the program. For example, suppose a serious drunk driving crash, in which a prominent citizen was killed, occurred during the course of the evaluation. It would be hard to say whether the responsible beverage server program or the emotional event and its publicity resulted in any reduction in impaired driving. For this reason, time-series evaluations lead to stronger conclusions when a program is introduced into different locations in different years, and trends over time are examined separately for each location.

Measurement → Measurement → Program → Measurement → Measurement

Figure 8.—Quasi-experimental design: time series.

Group 1: Measurement → Measurement → Program → Measurement → Measurement

Group 2: Measurement → Measurement → → → Measurement → Measurement

Figure 9.—Quasi-experimental design: time series with comparison group.

Time Series With Comparison Group

This design combines the use of a comparison group with a time-series design. Two groups or communities are compared over time to determine whether the experimental group changes more in the expected direction following the introduction of an intervention than a similar comparison group that does not receive the intervention. In the responsible beverage service example, impaired driving crash rates would be examined over time to detect general trends. The rates would then be compared following the implementation of the responsible beverage service program to determine whether the crash rates decreased in the experimental community more than would be expected from the general trend and more than in the comparison community. The combination of time series

and a comparison group allows a more confident conclusion that the program caused the observed effect because it controls for any general trends that might be occurring at the same time as the program. Figure 9 shows a time-series design with a comparison group.

Summary

This chapter detailed several types of experimental and quasi-experimental research designs that could be used in outcome evaluations. It also asserted that the strength of the research design rests in its appropriateness to the intervention it seeks to evaluate and its ability to attribute causality to the intervention. For more detailed information about research designs, consult the books listed in the appendix.

Chapter 4

Prerequisites to Data Collection

Once a research design to evaluate outcomes has been chosen, the evaluator-practitioner team must determine what data must be collected to answer the questions posed in the research. It is important that evaluators collect *only* data that can help answer evaluation questions, because data collection is often a burden on staff; it often interrupts normal work practices and sometimes requires staff assistance. Regardless of the types of data needed, and before any are collected, a number of issues must be dealt with.

Evaluations often involve original data collection (e.g., questionnaires, interviews, observations) and secondary data collection from established sources through archival research. Both types of data collection require securing the support and assistance of program staff, gaining access to records and people, securing consent to participate, ensuring confidentiality, and committing to provide useful feedback on a regular and timely basis. In addition, decisions must be made about how to choose individuals for inclusion in the data collection effort and how to cope with attrition. Each of these issues is detailed below.

Securing the Support and Assistance of Staff

Securing the support and assistance of intervention staff is a critical first step for both qualitative and quantitative research. It is part of the process of gaining the trust of staff who may perceive evaluation as a threat to the program and, hence, to their job security. Thus, people associated with the project (from administrators

to support staff and, if applicable, clients) should be well informed about the evaluation's purpose and activities. This may involve asking the evaluator to speak during a staff meeting and to provide a brief written description of what the evaluation is designed to assess, what types of data collection activities will be conducted, and the evaluation's projected timeline. Program administrators and evaluators might also develop a similar statement that could be distributed to clients, with clients being invited to direct any further questions about the evaluation to either program administrators or evaluators.

Although not all staff may have been involved in earlier collaboration with the program evaluator, staff service providers (who often tend to be most skeptical of evaluation research) should be invited to review the evaluation design and methodology. If possible, they should also be asked to provide suggestions as to how the methodology and planned logistical arrangements might be revised to ensure that the evaluation will be conducted in a way that is least disruptive to the day-to-day operation of the program.

If staff are expected to assist with data collection, it is important to assign roles and responsibilities and to record who must do what by when. If staff must collect information regularly, the usefulness of this information to the evaluation should be explained to them. Feedback on the progress of the evaluation, as well as preliminary findings, should be shared regularly with staff. Staff are more likely to invest themselves in the evaluation when they get regular feedback that relates directly to their roles in the project.

Gaining Access to Data

Approaching an organization for access to data that have already been collected requires a specific description of the information needed. Similarly, approaching an organization for access to its clients also requires a detailed statement. Many organizations have a formal application procedure that must be followed for access to data. For example, schools generally have processes for administering surveys, and police generally have procedures for releasing individual records for research purposes. Other

How To Gain Access:

- Write out evaluation questions and information needs.
- Find out about formal application processes.
- Concurrently, use informal process by cultivating an ally or group of insiders who can help.
- Maintain access with frequent reporting.

organizations require that requests for access be made in writing, sometimes supplemented with details about how the information will be used.

Gaining access to records and people may take several months or longer; evaluators and prevention administrators should begin negotiations early in the program to ensure adequate access to information. Personal introductions to recordkeepers—the individuals who maintain and are most familiar with the data—can be extremely helpful to the evaluators. Early on, the evaluator should explain to the recordkeepers what type of information is needed and for what purpose. These individuals often can point out useful data sources not already known to the evaluator.

Another approach is to convene a committee of representatives of different groups involved in the project. Members of the group may have the data needed for the evaluation, or they can assist in gaining access to the information. As the intervention proceeds, reporting information about activities and progress to collaborators helps maintain access and cooperation.

Securing Consent

Evaluators often require a signed, informed consent document from participants. This serves four purposes: (1) informing participants of the goals and procedures of the evaluation; (2) ensuring the confidentiality of responses; (3) informing participants of their right to decline participation in the evaluation, and that declining will not affect their receipt of program

How To Gain Consent and Ensure Confidentiality:

- Consult an institutional review board (IRB) that can assist in establishing safeguards. Call a local university or research firm and ask to speak to the chair of the IRB. Read the guidance on participant protection provided by the Office for Protection of Participants in Research.
- Train all data collectors to uphold the confidentiality procedures. Monitor them.
- Inform participants about the purpose of the evaluation and what will be asked of them. Assure them that confidentiality will be protected and that their participation is voluntary.
- Design consent forms and obtain participants' signatures.
- Plan to secure data through coding names, handling sensitive information carefully, and limiting access.
- Ensure that reports never link individual data with names.

services; and (4) describing the established procedures to safeguard individual confidentiality or anonymity. This document should be made available to the participants whenever requested. If children are involved, it is often necessary to get parents' consent in addition to the children's consent. Participants should always have an opportunity to ask questions about the evaluation, either in person or by telephone.

Ensuring Confidentiality

Substance abuse prevention evaluations often elicit sensitive information, such as alcohol and drug use or criminal behavior. One potential risk in any evaluation is the release of confidential information that could result in harm to the participant. Even less sensitive information, if publicly released or carelessly handled, could embarrass an individual and erode trust in the

evaluators. The privacy of individuals and the confidentiality of information must be maintained in all evaluations. Most institutions have institutional review boards (IRBs) to help guard the interests of evaluation participants. They weigh research procedures and their potential effects on individuals against the value of the information to be gained.

Evaluators and prevention administrators should always establish formal procedures to collect, store, and report confidential data. Some procedures are listed in the accompanying box. One recommended practice to help ensure confidentiality is to use identification numbers rather than names on project instruments. Another is to conduct interviews in a private area and to provide privacy when respondents are completing surveys. Evaluation data, especially sensitive information, should be stored in locked facilities. Evaluation reports should not link specific data, information, or responses with individual participants or their families.

Committing To Provide Regular Feedback

The importance of providing feedback to staff should not be underestimated. Evaluations often interrupt or interfere with their work, and staff frequently are not given feedback about the evaluation progress and preliminary findings. As a result, staff are often skeptical of, or even hostile toward, evaluation. Evaluation results, even if only preliminary, can be of great interest to staff. Sharing this information can do much to overcome staff resistance.

Before beginning the data collection process, evaluators should be prepared to communicate frequently with various audiences, such as intervention staff, clients, community groups, and the press. Early feedback will probably revolve around the data collection process itself (e.g., problems). Later on, evaluators might share preliminary findings based on an initial overview of a body of data as it is collected. Preliminary and more conclusive findings based on substantive analysis should be presented in the form of brief written updates, because program practitioners are not likely to have time to read detailed reports.

A basic update summary might include the following:

- A brief statement of the intervention's goals and objectives and the activity or activities designed to meet them;
- A summary of the activity or activities discussed in the update;
- The evaluation questions addressed in the update;
- A brief listing of research methods employed;
- A few graphs, charts, or bullet listings showing major findings; and
- A statement of what the results might mean to the audience.

Staff are an important first audience in that they can act as a reality check and give feedback on the accuracy of the findings. Staff also can use findings from the evaluation to improve activities. It is important to note that negative findings should not be concealed; rather, they should be presented and explained. Discussion of this sort is of significant value to a program that has accepted the evaluation as necessary and useful.

Sampling Strategies

If the number of people served by a prevention program is relatively small (fewer than 50), collecting data from the entire population can improve the credibility of the evaluation. However, many prevention programs cannot afford to study all the people they serve. For example, programs that provide prevention curricula to the general school population may be reaching tens of thousands of children each year. With such a large number of children served, the program will save time and money by studying a portion of the population served—a sample.

Determining the ideal size for a sample can be a complex process. Sample size determines the confidence intervals or margin of error in the statistics calculated from a sample—the larger the sample, the smaller the expected error. That is, the larger the sample, the more accurate (i.e., the closer to the value that would have been calculated for the entire population)

the statistics calculated for the total group will be, and the more powerful the study will be in detecting smaller differences. In most instances, programs will need help from a sampling expert to determine the best sample size.

Evaluation results from a sample may be safely generalized to an entire population, *but only if the sample is of a sufficient size and is representative of the total population*. For example, suppose a prevention program provides community workshops on setting family rules about drugs. Program staff want to know if their services change parental behavior. To answer this question, a broad cross-section of parents from a large portion of workshop locations (a representative sample) should be included in the study. A representative sample will show how well parents actually establish and maintain rules about drugs in their families after the workshops. A nonrepresentative sample (e.g., families who volunteer for an evaluation interview) will provide descriptive information about some parents but could give a biased picture of program effectiveness for all families.

Two probability sampling strategies used by experts to ensure a representative sample are simple random and stratified random sampling. In a simple random sample, each object (person, family, block, school, etc.) is drawn with equal probability and independently of every other object. Stratified sampling also involves random sampling, but probabilities of selection may be different for different groups, called strata. With simple random sampling, every unit in the population of interest has an equal chance of being included in the sample. With stratified random sampling, the population is first divided into strata (e.g., by ethnicity or gender); then simple random samples are drawn from each group or stratum. This strategy makes it possible to sample smaller groups with higher priority to ensure their adequate representation. Stratified sampling makes the analysis more complex and should always involve the guidance of a person competent in this practice.

The budget for evaluation is often the primary determinant of sample size. Frequently, information on intermediate effects and outcomes may be obtained only through costly personal interviews or survey questionnaires.

In determining the amount of information desired, the expense of increased sampling should be considered. There is generally a tradeoff between choosing a narrow and less expensive sample and a broader, more expensive sample.

Selecting Participants for Qualitative Evaluation

Selecting participants for qualitative research is as important as it is for quantitative sampling but may involve different procedures. Samples selected for quantitative studies are usually randomly chosen. However, since the goal of qualitative methods is to get the most information possible from smaller numbers of respondents, participants may be selected purposefully, although qualitative evaluators should be just as concerned about choosing a representative sample as are quantitative evaluators.

Qualitative evaluators can consider several strategies for sample selection. One method involves looking at typical cases only. Another method involves including extreme (unusual or special) examples to make for maximum variation in the sample. The extremes could vary by age, role, and experiences, and be compared on commonalities and differences among them or between the extreme and the typical examples. Critical or politically relevant cases could be studied to gain indepth information on specific situations. Knowledge of the population and setting through firsthand exposure to the setting or from information provided by program staff can help in selecting qualitative samples.

At times it may not be possible to get to know a setting well enough to select cases based on certain characteristics or differences. Snowball sampling, also called chain sampling, could be used in this situation. An evaluator starts with a few respondents and asks each of them to suggest other respondents who could provide useful information. For example, suppose a researcher wanted to interview high school seniors who have never experimented with alcohol or other drugs. He or she might ask a guidance counselor to suggest a few students to whom this description seems to apply. These students, in turn, might refer their friends

to the interviewer. The hazard in snowball sampling is that the sample may be homogeneous and unrepresentative, and thus be biased and ultimately inefficient, although convenient. Therefore, the decision to use snowball sampling should be made only after the focus of research is known.

Another sampling strategy is called convenience sampling. Participants are selected on the basis of easy access (e.g., selecting only participants who show up on a particular day). This method should be avoided except in exploratory studies because it may not meet the informational and analytical needs of an evaluation. The findings from this type of sample may be limited in applicability to other participants or locations.

Coping With Attrition

Sometimes people in a study cannot be found, or information cannot be collected from them. People move away, drop out of the program, refuse to complete a questionnaire, or are absent from school or work on the day of the study. Losing people from the study, called attrition, can affect results in two ways. First, if attrition is not considered when determining the desired sample size, it may result in too few people in the study, thus jeopardizing the confidence in the results. Second, people who cannot be located often share certain characteristics. Their exclusion jeopardizes the representativeness gained by random sampling or random assignment. For example, research has shown that rates of alcohol and drug abuse are higher among people who move frequently than among people who do not. Similarly, students who are frequently absent from school are more likely to abuse alcohol and drugs than students who attend school regularly.

Sample attrition, although potentially damaging to a study, can usually be addressed so the study is only partially compromised. One method is to select a larger sample for the study than is needed. If some participants drop out, the sample will remain large enough for analysis. Another technique to improve retention rates is to establish a tracking system to keep in contact with participants over the life of the study. Finally, pretest data, available for all par-

ticipants, can be used to assess the potential damage created by attrition. If attrition results in nonequivalent intervention and control groups, as might occur if the highest risk control group members are unavailable for the posttest but the highest risk intervention group members are successfully located, sound preintervention data can be used in an attempt to control statistically for some of the differences.

Summary

This chapter focused on critical practical issues that should be of concern to evaluators before they begin data collection. Whether implementing quantitative or qualitative data collection, evaluators should be concerned with establishing rapport with the prevention personnel and participants who are the focus of evaluation. Negotiating entry with these groups will facilitate cooperation and assistance when evaluators

are actually collecting data. Taking early steps to ensure access to archival data was also advised. Frequently, it can take several weeks to receive permission and retrieve documents from sources other than the program being evaluated. Evaluators should also secure a formal consent to participate in the study from those involved, particularly from program clients. Measures to guarantee that all data are kept confidential are also important. Another prerequisite is making a commitment to provide regular feedback to intervention staff throughout data collection and analysis. Decisions about sampling or selecting participants for data collection must be made before this process begins, and a strategy for coping with attrition must be adopted. While this information may not be new to the experienced evaluator, it should be reviewed by others who may not be familiar with conducting an evaluation.

Chapter 5

Common Data Collection Methods

Prevention programs are often complex, with multiple purposes, sites, and sources of data. Combining a variety of data-gathering strategies results in richer, more meaningful data. This chapter begins with a brief discussion of the distinction between qualitative and quantitative methods. Both quantitative and qualitative methods and data can be used to assess either intervention processes or outcomes. Table 5 illustrates how this is possible. There does tend to be a shift over the course of an evaluation, however, with more qualitative measures used early for process evaluation and quantitative measures used later for outcome evaluation. See figure 10 for a depiction of how this mixture tends to change.

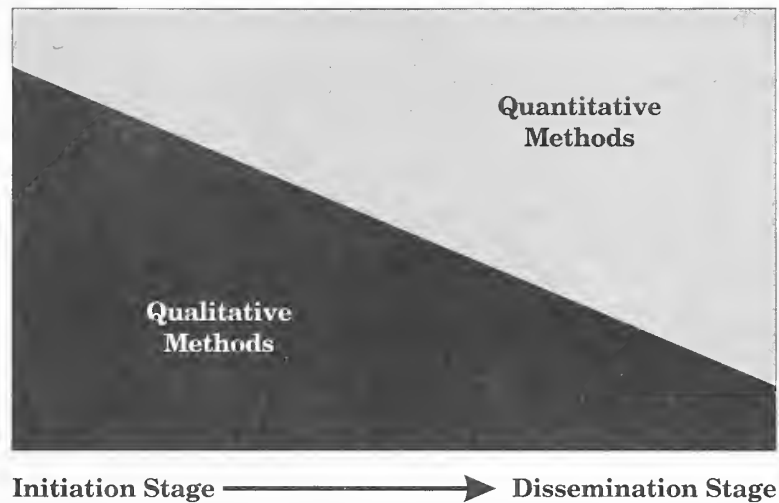
Qualitative Versus Quantitative Methods and Measurement

Qualitative methods result in descriptions of problems, behaviors, or events. Qualitative measures provide the stories that illustrate the nature of the problem addressed; the processes by which those problems were addressed; the complex, multifaceted dimensions of success; and the meaning of substance-related problem prevention for quality of life in a healthy community. Qualitative data provide more details or nuances but may be viewed as less objective and may be difficult to analyze or summarize and compare systematically.

In contrast, quantitative measurement consists of counts, rates, or other statistics that document the actual existence or absence of problems, behaviors, or occurrences. Quantitative data are generally considered more objec-

Table 5. Measures for Process and Outcome Evaluations

	Qualitative Measures	Quantitative Measures
Process Evaluation	<p>Focus group and interview data on support for intervention</p> <p>Observational data on mobilization efforts</p> <p>Interview data on organizational climate within the intervention</p> <p>Content analysis data on intervention's plan to produce desired program results</p> <p>Focus group and interview data on the nature of service delivery</p> <p>Archival data on the nature of service delivery</p> <p>Observational data on the nature of service delivery and context</p> <p>Interview data from program dropouts on causes of attrition</p> <p>Archival data on publicity, correspondence, recruitment efforts, attendance patterns, expenditures</p> <p>Content analysis data on media coverage</p>	<p>Questionnaire data on perceived need for the intervention</p> <p>Questionnaire data on nature and extent of problems</p> <p>Questionnaire data on participants' perceptions of intervention's "organizational climate"</p> <p>Archival data on numbers of people and constituencies involved in the intervention</p> <p>Questionnaire data on the nature of service delivery or policy implementation</p> <p>Questionnaire data from program dropouts to ascertain causes of attrition</p> <p>Questionnaire data on client satisfaction with program</p> <p>Counts of media coverage, enforcement activities, attendance at activities</p>
Outcome Evaluation	<p>Focus group data indicating changes in attitude and knowledge</p> <p>Interview data indicating changes in attitude and knowledge</p> <p>Observational data noting changes in the nature of the problem</p> <p>Archival data from intervention records concerning perceived effects</p> <p>Archival data on incidence of focal problem</p>	<p>Questionnaire data from participants and/or a control or comparison group from instruments designed to detect changes in attitude, knowledge, and behavior</p> <p>Followup questionnaire data reflecting changes in attitude, knowledge, and behavior</p> <p>Questionnaire data on client perceptions of program effects</p> <p>Questionnaire data on community perceptions of the focal problem and the intervention</p> <p>Observational data noting changes in problem incidence</p> <p>Archival data on intervention's monetary expenditures</p> <p>Hospital and police records on incidence of focal problem (e.g., injury, death, vandalism)</p> <p>Counts of alcohol and tobacco purchase attempts from licensed establishments</p> <p>Police enforcement records</p> <p>Counts of alcohol and tobacco nonmedia advertising</p> <p>Cost-benefit data on monetary expenditures and documented positive effects</p>



Stages of Program Development

Figure 10.—Research methods employed in comprehensive program evaluation.

tive and easier to summarize and compare than qualitative data, but they may not provide all the information needed for interpretations of findings.

Qualitative Methods

Qualitative methods produce findings that show patterns or inconsistencies among data. Frequently employed qualitative data collection methods include interviews, focus groups, participant-observation, and archival research. Each of these methods is described below.

Interviews

Interviews are a common qualitative data collection technique. There are three standard interview formats: structured, semistructured, and unstructured. Structured interviews (face-to-face questionnaires) are generally considered a quantitative method and are discussed later in this chapter. Other types of interviews that allow for open-ended responses provide a wealth of data about a specific topic but require additional time to analyze. Basically, the less structure is applied in the interview, the more the information gained will depend on the abilities of the interviewer. Interviewers must be skilled in interviewing techniques, have excellent interpersonal skills, and be sensitive to

cultural differences. Usually, interviews should occur in a private setting, with assurances of anonymity and confidentiality.

Semistructured Interviews. Semistructured interviews involve asking many respondents the same series of questions; however, responses are not limited to a given set of answers. Consequently, the evaluator is assured that data are collected in response to the same series of questions from a number of individuals. Both semistructured and unstructured interviews create an interactive situation that frequently involves the interviewer's asking additional questions to draw out detail. Many observers believe that semistructured and unstructured interviews facilitate more candid responses. Because the content of responses may range tremendously, however, significantly more time may be required for data analysis.

Unstructured Interviews. Unstructured interviews may begin with a series of questions that the evaluator wants answered, but the interviewer tends to let the conversation flow naturally rather than being constrained by a set outline. What is paramount in the unstructured interview is that the respondent feel free to answer questions in his or her own words and that the evaluator be prepared to ask unantic-

pated questions based on information contained in responses to earlier questions. Unstructured interviews differ from conversations in that the evaluator is primarily concerned with gathering information and not with contributing to conversation—that is, extended participation by the evaluator is only a means of clarifying question intent and steering the interview.

Unstructured interviews are useful when exploring sensitive issues, emerging events, or unique experiences. They are also common when a respondent is expected to provide unique information. Consequently, the unstructured interview method is not usually employed when an evaluator is trying to collect information about the same topics from several persons.

Focus Groups

Focus groups are informal discussion sessions that typically involve 6 to 10 people. Discussions center around specific topics and usually last between 1 and 2 hours. Participants with similar specific characteristics are invited to attend. For example, if one suspected that adolescent substance abuse was particularly pervasive among male athletes in a small town and wanted to know more about this phenomenon, one could conduct a focus group with adolescent male athletes who admit to engaging in alcohol or drug use. Or, one could conduct a focus group with junior and senior high school coaches of male sports teams to explore why they think the problem exists. Or, one could conduct a focus group with parents of adolescent male athletes to assess their perspectives on the problem.

Facilitators of focus groups use a discussion guide with topics or questions to be covered. The groups are frequently taped with permission from participants. Generally, a skilled facilitator guides the group interaction, while a colleague takes notes, runs the tape recorder, and handles other logistics. Focus groups differ from interviews in that the elements of interaction and discussion among respondents are added. Moreover, it is believed that because participants share certain characteristics, they will be inclined to express their opinions honestly during the discussion.

Participant-Observation

Participant-observation differs from standard observation in that the evaluator participates in the activity being observed. The evaluator usually takes notes on the progress of the activity, the physical setting, patterns of interaction and decisionmaking, and responses to planned and unplanned occurrences. As a participant, the evaluator also records his or her own reactions to the nature of the activity, noting, for example, if the activity was stimulating or tedious. In programs serving marginalized groups (e.g., recovering substance abusers, the disabled, the elderly), observers in general may not be welcome, but participant-observers who resemble group members to some degree may be welcome.

Archival Research

Archival research with a qualitative focus involves examining written records in order to understand a program better. Meeting minutes, journals, logs, program and agency correspondence, and other historical documents inform evaluators about program operations. These materials also generate ideas for other questions to pursue in interviews and observations.

Quantitative Methods

By definition, quantitative methods produce measurable findings that are expressed in numbers, such as amounts, ratios, and percentages. Frequently employed quantitative data collection methods include a variety of survey questionnaires, observation, and archival research. These methods are described below.

Questionnaires

Questionnaires are lists of questions that ask about a range of behaviors and/or opinions. Questionnaires generally restrict answers to those provided on the form. That is, little or no space is provided for respondents to answer questions in a way that diverges from the response range conceived by the questionnaire's creator. Questionnaires can be designed for a specific program, intervention, or population. In some cases, standard questionnaires used in other research and evaluation efforts may be more appropriate.

For easy statistical analysis, each possible response is assigned a numerical value. In turn, this value is used to code actual responses; thus, questionnaire data can be processed and analyzed quickly.

Ideally, questionnaire surveys should be brief, while still collecting the necessary data. A new instrument should be pilot-tested—that is, tested for clarity, readability, and suitability with a smaller but similar population—before it is administered in an evaluation. Based on feedback from the pilot test, revisions to the instrument may be necessary. Good questionnaires are worded in an accessible language and at an appropriate reading level. They are also designed and implemented in a culturally sensitive fashion. Questionnaires are usually administered in one of three ways.

Self-Administered Questionnaires

Self-administered (or written) questionnaires require that the respondent complete all portions of the form without assistance. Self-administered questionnaires can be distributed directly by the evaluator to the respondent or indirectly by mail or other third party (such as a classroom teacher or an employer). When questionnaires are distributed directly, more completed questionnaires are likely to be returned.

One drawback of self-administered questionnaires is that not all respondents may understand or interpret the questions and give responses in the way that the evaluator intended. Consequently, the questionnaire may produce data that are not completely accurate or, in some cases, are contradictory. Another problem results when respondents do not answer all questions. In other cases respondents may not believe that any of the given responses accurately or even remotely express their particular viewpoint. In these cases, they may choose more than one response or choose not to respond at all, thereby invalidating their response to the question and contributing to different response rates for each question on the form. Finally, self-administered questionnaires are subject to selection bias that can affect findings. For example, those who liked the program may be more likely to return questionnaires than those who did not. Mail surveys are well known for producing relatively low response rates.

Face-to-Face Questionnaires (Structured Interviews)

Face-to-face questionnaires, or structured interviews, require that an interviewer ask all the questions on the form directly to the respondent and record responses on the questionnaire form. Under these circumstances, the respondent can often ask for clarification if questions are confusing. Some believe that the one-on-one nature of structured interviews increases the likelihood that all responses will be completed because the interviewer is not likely to let a question remain unanswered. Structured interviews also allow for easier comparisons among respondents than semistructured interviews (discussed above). On the other hand, respondents may be less likely to answer sensitive questions accurately in a face-to-face setting. Structured interviews are also more costly to implement because they require the use and training of interviewers.

Telephone Questionnaires

Telephone questionnaires are also surveys that require the use of an interviewer, and, like face-to-face questionnaires, the involvement of this person allows for clarification of questions that seem confusing to the respondent. Moreover, telephone questionnaires yield nearly as high a response rate as structured interviews but at a much lower cost, especially if the respondents are spread over a large geographic area. Confidentiality may be more difficult to maintain in telephone interviews, however, because the interviewer cannot completely control for privacy.

Observation

Observation is another method employed by evaluators who wish to collect data in a natural setting. Observers should be as unobtrusive as possible as they record behavior using checklist instruments that count the frequency of certain behaviors or interactions. Checklists are relatively easy to complete. The observer looks for certain predetermined behaviors and checks them off when they occur. For example, the observer may watch to see if salesclerks ask for age identification before selling tobacco or alcohol to young customers.

Interactional instruments measure the dynamics of human interaction. They note behaviors at certain times and record patterns

of interaction, such as who talked to whom. An interactional instrument may be suitable to measure certain prevention objectives. For example, if the intent of a program is to increase friendship attachment cliques in a middle school, observers could record the patterns of interactions among students in a classroom.

Regardless of the data source, data collectors must be selected carefully and trained to use consistent techniques that do not introduce bias into the data. Bias can occur when data are not collected systematically from all the targeted groups or individuals, when differences among observers produce differences in data collected, or when data collection itself introduces systematic influences on what is being observed.

Archival Research

Archival research is one of the most frequently used and least expensive data collection methods. Information that is normally kept as part of agency or organization operations (e.g., school records, health care and social services records, police and juvenile court records) can provide important statistical data for evaluations and can supplement the information gathered from other data sources.

In addition, major data sets indicating trends in specific substance abuse-related problems may provide information relevant to the evaluation or useful in comparisons of local statistics with county, State, and National statistics. State and local alcohol beverage control agencies systematically collect data on active liquor licenses. Through its Alcohol Epidemiologic Data System, the National Institute on Alcohol Abuse and Alcoholism maintains data on alcohol consumption in the 50 States and the District of Columbia. State health departments maintain records on acute hospital discharges and deaths attributable to alcohol-, drug-, and tobacco-related illnesses, as well as records on substance abuse-related health problems in newborns. State departments of social services often contain records on substance abuse-related child abuse and neglect. The National Highway Traffic Safety Administration's Fatal Accident Reporting System tracks alcohol-related traffic fatalities. SAMHSA's Uniform Facility

Data Set (UFDS) (formerly NDATUS) tracks treatment facility admissions and financing.

Two limitations on the use of agency archival data for evaluation purposes are confidentiality (written, informed consent of participants is sometimes necessary) and limited relevance (needed data are not included or are incomplete or inconsistently collected).

Table 6 summarizes the data collection methods discussed in this chapter, noting the strengths and weaknesses of each.

Summary

This chapter described commonly used quantitative and qualitative data collection methods, noting the appropriate uses and strengths and weaknesses of each method. Both quantitative and qualitative research methods should be employed to assess program functioning as well as program outcomes. Although collecting a wide array of evaluative information may at first appear daunting, measurement development and data collection do not have to be the work of just one person, group, or agency. Rather, the evaluation can be a coordinated effort of individuals and organizations.

Data can come from different sources. Some information may be available from record archives. For example, law enforcement agencies maintain records of traffic crashes, and hospitals maintain records of the occurrence of various injuries and diseases related to substance use. Other information may be collected through measurement instruments that are already routinely administered (e.g., school surveys administered by departments of education).

Data collection methods can also complement each other. For example, observation of a classroom can indicate how well children appear to respond to the program, and review of their school records can help determine the impact of the program on academic performance. Good evaluations generally employ several types of data collection methods so that a more complete and accurate overview of intervention operations and effectiveness can be documented and defended by various data sources.

Table 6. Common Data Collection Methods

Qualitative		
Method	Strengths	Weaknesses
Semi-structured Interview	<p>Standardized questions and steering of the interviewer tend to ensure that data are collected in response to the same series of questions from a number of individuals.</p> <p>Open-ended responses result in significantly more in-depth data than structured interviews with standardized responses.</p> <p>The somewhat conversational nature of this method may promote more candid responses.</p>	<p>More time consuming than structured interviews.</p> <p>The range of content responses requires considerably more time for analysis than for structured interviews.</p> <p>Poor interviewing or interpersonal skills of the evaluator may result in the introduction of bias (e.g., leading respondents).</p>
Unstructured Interview	<p>Open-ended responses result in tremendously detailed data.</p> <p>Often implemented in tandem with participant-observation, and because of the highly conversational nature of this method, interviews result in quite candid responses because the respondent feels very comfortable and familiar with the interviewer/evaluator.</p>	<p>Usually quite time consuming to conduct and to analyze data.</p> <p>Appropriate only when seeking critical and detailed information that can only be, or best be, gathered from one source.</p>
Focus Group	<p>Extended nature of discussion lends itself to group's exploration of specified topics in-depth.</p> <p>Because of similar characteristics among group members, participants tend to share opinions and feelings more honestly.</p>	<p>Analysis of discussion content as well as behavior can be time consuming.</p>
Participant-Observation	<p>Yields considerably more detailed information about program activity than conventional observation.</p> <p>Because the participant-observer looks like other participants, his/her presence may have less effect on the normal behavior of other participants than would the presence of an outside observer.</p>	<p>Transcription and analysis of field notes is time consuming.</p> <p>Among marginalized groups, participant-observers who do not share many characteristics of the study population may encounter difficulty being accepted.</p>

Table 6. Common Data Collection Methods (continued)

Qualitative		
Method	Strengths	Weaknesses
Archival Research	<p>Provides historical information about the study subject.</p> <p>One of the least costly evaluation methods.</p> <p>Information may generate ideas for better questions to pursue through interviews and to explore through observation/participant-observation.</p>	<p>Access to some types of institutionalized records may be difficult to obtain.</p> <p>Data quality may be variable.</p>
Quantitative		
Method	Strengths	Weaknesses
Self-Administered Questionnaire	<p>Collects a great deal of information from a large number of individuals in a standardized, and therefore easy-to-analyze, way.</p> <p>May be implemented in a relatively efficient and inexpensive way.</p>	<p>If not worded well or at an appropriate reading level, will cause confusion in the respondent and possibly lead to erroneous data.</p> <p>Respondents may not answer all questions.</p> <p>Frustrated by limited response choices, respondents may provide multiple answers to questions and nullify their responses.</p> <p>Questionnaires distributed by mail are subject to selection bias and generally result in lower response rates than for other types of questionnaires.</p>
Face-to-Face Questionnaire (Structured Interview)	<p>Collects a great deal of information from a large number of individuals in a standardized, and therefore easy-to-analyze, way.</p> <p>Presence of interviewer tends to ensure that all questions are answered in the desired format.</p> <p>Presence of interviewer allows for clarification of question meaning.</p> <p>Data may be analyzed relatively inexpensively.</p>	<p>Respondents may be less inclined to answer sensitive questions truthfully in a face-to-face format.</p> <p>More costly and more time consuming than self-administered questionnaires.</p> <p>Greater risks of loss of confidentiality.</p>

Table 6. Common Data Collection Methods (continued)

Quantitative		
Method	Strengths	Weaknesses
Telephone Questionnaire	<p>Collects a great deal of information from a large number of individuals in a standardized, and therefore easy-to-analyze, way.</p> <p>Involvement of interviewer tends to ensure that all questions are answered in the desired format.</p> <p>Physical distance between interviewer and respondent may increase likelihood for honest responses to sensitive questions.</p> <p>Data may be analyzed relatively inexpensively.</p> <p>May be more cost efficient than other survey strategies if a large geographic area is targeted.</p>	<p>More time consuming than self-administered questionnaires.</p> <p>Significantly more costly than mail surveys.</p> <p>Confidentiality may be more difficult to maintain because the interviewer cannot control for privacy on the respondent's end of the telephone line.</p>
Observation	<p>Once access to setting has been secured, data may be collected easily and unobtrusively.</p> <p>Counting and use of checklists make data analysis relatively easy and inexpensive.</p>	<p>Knowledge of observer's presence may alter normal actions and behaviors of the individuals being observed.</p> <p>Observer's presence may not be welcome by activity's participants.</p>
Archival Research	<p>Provides systematically collected historical information about the study subject.</p> <p>One of the least costly evaluation methods.</p> <p>May generate other questions for new ideas to pursue about the study topic.</p>	<p>Access to some types of institutional records may be difficult to obtain.</p> <p>Records may not contain relevant data, or data may be incomplete or inconsistent.</p>

Chapter 6

Concepts in Data Analysis

Data analysis is a process that involves categorizing, ordering, manipulating, and summarizing data. Throughout this process, evaluators draw preliminary conclusions from the data and seek verification of these conclusions from various sources within the database. This chapter discusses basic concepts in quantitative and qualitative data analyses that frequently appear in evaluation reports and with which prevention planners should be familiar. It also addresses which types of data analyses are appropriate at each stage of program development. This chapter does not detail specific data analysis methods (e.g., regression analysis, path analysis). Such information is available in a variety of methods handbooks and is appropriate reading for evaluators who must actually carry out data analysis.

Every evaluation plan should include a data analysis plan that specifies how the evaluator intends to summarize the collected data. There are two types of quantitative data analysis that should be included in any comprehensive data analysis plan: descriptive analysis and relational analysis.

Approaches to Quantitative Data: Descriptive Analysis

Descriptive analysis summarizes information in order to provide an initial picture of the prevention program. Descriptive analysis provides basic information about the variables included in the study. For example, descriptive analysis may try to answer the questions, "How many pregnant women in our prenatal program reported smoking tobacco in the year preceding

their pregnancy?" or "How many employees are affected by a new smoke-free workplace policy?"

Counting and simple arithmetic are sufficient for carrying out descriptive analysis. The evaluator of an individual-oriented intervention can determine such things as how many people were included in the study; their breakdown by age, sex, or any other characteristic considered important for the study; the types of interventions to which they were exposed; how often they attended sessions; and the number who reported specific behavior relevant to the study. The evaluator of a policy intervention might explore how many people were exposed to the intervention, how long the intervention lasted, and how consistently the intervention was applied. Once this information is gathered, it can be presented either as raw numbers or as percentages of a larger number or the larger study subject pool.

Measures of Central Tendency

After basic descriptive statistics are compiled, evaluators may wish to determine central tendency. In evaluation research, central tendency is often presented in terms of the mean, median, and mode. A mean is simply an average, and it can be calculated by dividing the sum of values by the number of values. A median is the point at which 50 percent of the values fall below it and 50 percent exceed it. The median is often presented along with the mean because the mean is influenced by the high and low extremes of a range. The mode refers to the value most often given by respondents. Table 7 illustrates the mean, median, and mode when 13 participants in a prevention workshop were asked to rate the overall course on a scale of 1 to 10.

Standard Deviation

It is often useful to know how much variation exists among the scores. For example, suppose the evaluator of the prevention workshop wanted to know whether participants rated the workshop fairly consistently or whether there

Table 7. Sample Rating Scores on Scale of 1 (Extremely Poor) to 10 (Excellent)

8	
9	
5	
8	Mean = 7.8 (98/13)
8	
10	
7	
8	Median = 8
8	
7	
8	
7	
8	Mode = 8

was a great range of opinion. In order to determine the degree to which responses were consistent, it is necessary to calculate the standard deviation.

Standard deviation refers to the spread of scores away from the mean score, noting where the bulk of responses lie. Applying the formula for computing standard deviation to the data from the prevention workshop example, we learn that the standard deviation value is 1.17. Given that certain statistical assumptions hold true, this means that 68 percent of the rating scores fell within one standard deviation above and below the mean or between 6.63 and 8.97. Since we know that the 10-point rating scale used to evaluate the workshop used increments of one, we can also say that 68 percent of the respondents gave the workshop a rating of either 7 or 8. In light of the mean (7.8), one can conclude that the data contained a very low standard deviation and that participants rated the workshop fairly consistently.

Consistency of ratings (a low standard deviation) is often a welcome sight to evaluators, regardless of whether ratings are consistently good or bad. A low standard deviation suggests that later conclusions based on complex statistical (relational) analysis are more easily defensible.

Stages of Development and Descriptive Analysis

The simple yet informative nature of descriptive analysis suggests that it is best used as a backdrop to relational analysis in all experimental and many quasi-experimental research efforts. It is also quite well suited to quasi-experimental research designs to provide general descriptive information about a program itself and not its effects. Moreover, in light of the multifaceted nature of the most complete evaluations, descriptive analysis may be the primary type of analysis on data collected during the early stages of an intervention.

During the Initiation and Planning Stages, staff are likely to be concerned with determining the nature and extent of problems and assessing the community's level of interest in perceived problems as well as any effort to address them. Prevention planners may also wish to know the perceived causes of the problem as well as the perceived obstacles to overcoming it. This type of information can be gathered through quantitative and/or qualitative research instruments. If surveys are employed, descriptive analysis is probably the only type of quantitative analysis that one might expect to perform on the data. This analysis would provide information about which segments of the community are concerned about the problem (e.g., men, women, parents, business people, the elderly) and about how concern varies among different constituencies. It could also provide information about how particular constituencies interpret causes and solutions. These causes and solutions could be rank ordered according to the frequency with which they were cited.

During a Pilot-Testing Stage, evaluators may wish to note participants' level of satisfaction with the intervention, as well as to document its actual implementation. Once again, quantitative or qualitative data collection techniques can be used. If the evaluation plan calls for a survey of participants or observation with the use of checklists and other instruments to record interaction, descriptive analysis will be necessary to determine how participants evaluated the program (e.g., how many participants rated the intervention unsatisfactory). Descriptive

analysis will also indicate whether planned occurrences and desired behaviors actually occurred and how often they occurred within the context of the intervention.

In the Implementation and Stabilization Stages, tasks and research questions tend to be somewhat similar in that questions regarding adequacy of the intervention, obstacles, costs, and effects are ongoing concerns. These are also stages in which outcome-related questions become more urgent. Consequently, descriptive analysis becomes essential for providing a framework for understanding the results of more complex statistical analyses. In addition, descriptive analysis remains necessary to address such concerns as participant satisfaction and other support for the intervention.

The Dissemination Stage involves, in part, assessing projected needs. It may also involve expanding the intervention to either implement new interventions or institute the same services in a new area. If the intervention expands, evaluators will return to some of the questions addressed in the Initiation and Planning Stages and employ descriptive analysis accordingly.

Approaches to Quantitative Data: Relational Analysis

The second type of quantitative analysis that one would expect to find in a data analysis plan is relational analysis. Relational analysis allows one to understand the relationship between key variables in the study or to test a hypothesis. It is particularly well suited to outcome evaluations. For example, relational analysis could determine whether requiring a beverage server training course was related to fewer alcohol-related traffic accidents in a given community. Relational analysis involves making comparisons between people who were subjected to the intervention and those who were not; in other words, analysis must be based on data from an experimental design or a quasi-experimental design that controlled for spurious effects of variables other than the intervention.

Statistical Significance

The results of relational analysis are presented in the form of statistics, highlighting what are

Table 8. Contingency Table (Example)

	Employees Attending Program A N = 160	Employees Attending Program B N = 190	TOTALS
Employees who requested information	60 (37.5%)	25 (13.2%)	85 (24.3%)
Employees who did not request information	100 (62.5%)	165 (86.8%)	265 (75.7%)
TOTALS	160 (100%)	190 (100%)	350 (100%)

termed statistically significant results. Statistical significance is another way of saying that a particular relationship between variables occurs so frequently in the data that the relationship's existence can probably not be attributed to chance. Statistical significance is determined through one of several tests. Generally speaking, statistical results are said to be significant when relationships between variables reach the .05 (or in some cases .01) level—that is, when a relationship has a less than 5-percent probability of happening by chance.

When considering statistically significant findings, prevention planners should be aware of certain limitations of the term. It simply points out that a statistically significant relationship exists. It does not mean that an intervention caused the relationship. Such a determination depends on the research design and more complex statistical analyses. In addition, statistical significance does not indicate that the finding is strong enough to be important to program planners and policymakers. This determination should be based on an analysis of the costs and benefits of the intervention, among other issues.

Contingency Tables

A relatively simple way of comparing relationships between or among variables is through the creation of a contingency table. A contin-

gency table presents the number and percentages of participants in at least two groups according to at least two variables. It is useful when variables are clear-cut categories, like sex or race, and not when they can be rank ordered, like age.

Table 8 is a contingency table for participants in two different employee assistance programs who requested followup literature.

This table shows that people who attended Program A requested information more often than people who attended Program B. If one wanted to determine whether the difference between the two groups is statistically significant, one would have to employ a test of statistical significance.

Means Comparisons

One way of comparing relationships between variables when they can be rank ordered is by comparing the means of different groups. For example, suppose one wanted to know if a 4-week, parent-oriented substance abuse prevention program affected how many times parents discussed the topics of alcohol, tobacco, and illicit drugs with their children in the 6-month period following completion of the program. One would compare the mean number of discussions for a sample of parents who attended the program with the mean number of dis-

cussions for the control group of parents who did not attend the program. Once again, if a difference was detected, it would have to be tested to determine if it is statistically significant.

Means comparisons can also be done for three or more groups. Returning to our example, let us say the three groups were parents who attended three or four sessions, parents who attended one or two sessions, and parents in the control group. Means comparison can be done for more than two groups while controlling for another measure (say, whether it was a single-parent household). Both of these approaches would require a different test of statistical significance for any difference that might be detected.

Time-Series Analysis

As discussed in chapter 3, a time-series research design is one way in which the effects of an intervention can be assessed. The type of statistical analysis used in time-series designs involves examining data collected before, during (if appropriate), and after the intervention. If, for example, one wanted to know if a temporary local ordinance banning cigarette vending machines resulted in decreased cigarette sales, one might track cigarette sale data for several months before the enactment of the ordinance, the period during which the ordinance was in effect, and several months following the suspension of the ordinance. A sustained difference in sales during—and only during—the period of intervention allows for a plausible assertion of causality. Statistical techniques are used to determine whether observed changes over time are statistically significant.

If the collection of data during the intervention (a 1-week prevention education course, for example) seems impractical, data could then be collected for an extended period of time before, as well as after, the intervention.

Correlational Approaches

In evaluation research, regression analyses are quite common. These types of statistical analyses are more complex than those previously discussed, and they focus on the relationships between variables that can be assigned value.

Correlational or regression approaches indicate whether one variable increases when another variable either increases or decreases—for example, whether the increase in a student's exposure to antialcohol messages in school results in a decrease in the incidence of alcohol consumption.

Regression analysis should be performed by a trained evaluator. It should also be noted that the results of regression analysis can indicate whether a positive or negative correlation exists between variables and whether a statistically significant relationship exists between variables. The results, in and of themselves, cannot support a conclusion that an intervention caused the relationship.

Advanced Analyses

There are other forms of analysis that can be applied in the evaluation of intervention outcomes (e.g., analysis of variance, analysis of covariance). Like correlational or regression analysis, they should be performed by an evaluator trained to do so.

Multiple Approaches to Data Analysis

Clearly, if one intends to determine the effects of an intervention on a group, relational analysis is required. Specifically, this analysis should center around data collected at the start of the implementation stage through the stabilization stage (and possibly afterward), when evaluation questions focus on outcomes.

In light of limited funding for evaluation and, perhaps, the limited availability of experienced program evaluators, it is reasonable to ask what types of data analysis should be employed. As has been stated earlier in this guide, multifaceted research designs requiring multiple approaches to data analysis work best. Every program can carry out useful descriptive analyses of quantitative data. With the guidance of an experienced researcher, means comparisons and contingency tables (two elementary forms of relational analysis) are certainly possible. More sophisticated relational analyses require considerable training and experience.

Statistical analysis of quantitative data is well and good, but it uses only part of the evaluator's collection of useful tools. Regardless of whether an evaluation plan and budget allow for rudimentary or highly complex statistical analysis, the collection and analysis of qualitative data should be included in any data analysis plan.

Approaches to Qualitative Data Analysis

Qualitative data contain verbal descriptions of problems, behaviors, or events rather than counts or rates of events or occurrences. Qualitative data provide detail, context, and qualifying considerations that relate to the problem being addressed, and to the process of how the problem was addressed.

Qualitative evaluators are among the first to assert that qualitative data are "very messy." But, even though qualitative data in their initial form tend to be verbose and for the most part unstructured, qualitative research remains one of the most valued components in any evaluation. Unlike quantitative approaches, which begin with sharply defined categories of study and measure the degrees to which they exist, qualitative methods begin with categories but anticipate that the meanings behind them may vary according to how they are used, as well as who uses them. It is the task of the qualitative evaluator to discern patterns and differences in meaning among the variety of perspectives and to hypothesize about the causes of these patterns and differences.

For example, a survey may reveal that the vast majority of the respondents believe that young people use alcohol and illicit drugs because they have no attractive alternative choices. Through qualitative research, one can begin to understand why people in general or segments of the population (e.g., whites, blacks, others) believe this to be true. If there is variation among responses, the qualitative evaluator will attempt to find out its causes. Patterns discerned from interview data, for example, should then be compared with data derived from other sources, such as participant-observation and focus groups. Ultimately, the analysis may reveal contextual issues that could signifi-

cantly affect the development of any intervention plan as well as explain the success or failure of intervention effects.

The remainder of this chapter briefly describes the basic progression of qualitative data analysis in most evaluation studies: data transcription, data reduction, and conclusion drawing.

Data Transcription

The rawest forms of data are field notes generated through participant-observation. These notes contain the evaluator's immediate and initial thoughts about occurrences and should be elaborated upon further in writing. For example, if a program evaluator attends a series of meetings in which community members draft an ordinance to restrict alcohol use in public settings, field notes should be transcribed after each meeting. The more time that elapses between taking the field notes and writing them up, the greater the likelihood that the evaluator will not recall the full meaning behind the notes. Qualitative data from interviews, focus groups, and archival research should be treated in a similar manner.

Data Reduction

Once all recorded data are transcribed, they and other data are organized into more user-friendly formats through the use of codes. Qualitative researchers often begin data collection with general code categories based on research questions. After initial review of collected data (through data transcription), these general codes are refined in order to facilitate the discernment of patterns and inconsistencies. These revised codes are then assigned to "chunks" of transcribed data. This is a critical first step in qualitative data analysis because it involves simplifying what appears initially in a very unstructured or "messy" form.

For example, suppose one evaluation question concerned the perceived causes of substance abuse-related problems. A general code category for collected data might be called "causes of problems." After initial review of data, this general code might be refined to contain several subcode categories, such as "avail-

ability of substances" and "unemployment." These codes and subcodes are then assigned to data. Sometimes it may be relatively easy to assign codes, as when respondents in semistructured interviews are asked to comment on causes. Other times, pertinent information may be embedded in other types of commentary, as when focus group members discuss solutions to substance-related crime (e.g., visible police presence). In such instances, data "chunks" may be double coded. Multiple coding of qualitative data is fairly common.

Data Display

Computers have greatly facilitated the process of assigning codes to bits of data. If possible, a body of data that has been transcribed with common word-processing software can be easily adapted into a "code and retrieve" software program, such as The Ethnograph. This type of program is also capable of displaying all data chunks that are associated with a particular code or subcode. Once the evaluator is able to view qualitative data in this manner (either with or without the assistance of computer software), the final process in qualitative data analysis (conclusion drawing) becomes possible.

Drawing Conclusions

Preliminary conclusions are sometimes used as a basis from which the evaluator may seek answers to a limited number of additional questions that are quite focused. For example, if data from several research activities suggest that the nature of law enforcement contributes to the pervasiveness of substance abuse-related problems, then the evaluator may seek interview data from one or two key individuals (e.g., a law enforcement official or beat officer) whose knowledge about the topic can speak directly to the plausibility of the preliminary conclusion. An evaluator might also conduct additional archival research on such objective measures as the number of assigned officers, response times, and basic crime statistics for the target area as compared with a larger area (e.g., town, city, county). On the basis of such information, a preliminary conclusion may be asserted more definitively, or it may simply be presented as a popular perception that is only somewhat sub-

stantiated, or that perhaps is refuted by other sources.

The process of verifying preliminary conclusions may not always require a return to data collection. Sometimes, a program evaluator might share preliminary conclusions with program staff or original data sources and request feedback. Or an evaluator may simply return to the original raw or coded data. In any event, verification of preliminary conclusions is essential in any qualitative data analysis effort. If it is not performed and the process not documented in the final evaluation report, conclusions will be open to sharp criticism—justifiably.

Summary

This chapter reviewed concepts in quantitative and qualitative data analysis that should be understood by all prevention planners. Although many program budgets may not allow for sophisticated quantitative analyses, every program can still conduct useful forms of descriptive analysis, as well as elementary forms of relational analysis on quantitative data.

This chapter also asserted that different types of analysis are better suited to certain types of data (i.e., data collected at specific stages in a program's development). Descriptive analysis of quantitative data, as well as most qualitative analysis, is likely to be performed on data collected during the early phases of an intervention's existence (Initiation, Planning, Pilot Testing). Relational analysis that is oriented toward assessing outcomes is better suited to data from later stages (Implementation, Stabilization, and Dissemination).

Since most prevention interventions are likely to be involved in conducting at least some degree of qualitative research during the course of an evaluation, this chapter addressed the basic progression of qualitative analysis. If participant-observation or focus groups were included in the evaluation plan, data should be transcribed. Data should then be coded and analyzed for patterns and inconsistencies. The initial or preliminary conclusions drawn from qualitative data should be verified in some fashion by the evaluator in order to produce credible final conclusions.

Chapter 7

Conclusions and Recommendations

Carrying out a credible and useful evaluation is not easy. Local entities responsible for delivering services or implementing policy generally do not employ in-house evaluation staff. Spending scarce resources to purchase evaluation services is a difficult choice for program administrators who may not realize just how important evaluation is to developing effective prevention interventions and ensuring the continuation of these efforts. But the need for demonstrating the value of prevention has never been so critical to the future of prevention programming as it is today. This guide was designed to assist collaborative evaluator-practitioner teams in designing and implementing useful evaluations.

Collaboration among evaluator, prevention administrator, and intervention staff is essential in any evaluation effort. Such collaborations are much more likely to ensure the development and implementation of appropriate, comprehensive, and useful evaluations than traditional researcher-driven evaluations. At a time when decreased funding for prevention has spurred calls to document intervention effectiveness, the ability to produce useful, high-quality evaluations has never been more important.

Critical to designing useful evaluations is beginning the process with a recognition of the type of intervention approach. This information will influence not only how the evaluation should be designed but also what types of data should be collected. Another important element in any evaluation is early attention to an intervention's stage of development. Attempting to assess intervention outcomes during the early

stages of implementation is not only inappropriate, it is also wasteful of precious dollars. Process evaluations that document how an intervention functions before it achieves outcomes in later stages are much more appropriate and provide invaluable information.

Any formal evaluation plan begins with the study design. Given that many prevention efforts are multifaceted in nature, the use of evaluation plans that assess intervention processes as well as outcomes through a variety of approaches is advised. This involves incorporating either experimental or quasi-experimental research designs in order to be able to make assertions about intervention outcomes.

Before data collection begins, several practical issues should be addressed. Among the most common are securing the support and assistance of staff, access to data, and the consent of participants, as well as ensuring confidentiality to participants and guaranteeing that collected data—often quite sensitive—is maintained securely. Another practical issue concerns providing regular feedback on the evaluation's progress and preliminary findings to staff and other constituencies. Many evaluators fail to give such feedback, and this is the principal reason why many practitioners remain skeptical, or even hostile, toward evaluation. The evaluation process is often an intrusion into the work of the staff, particularly when they have been asked to assist with data collection. Yet they also believe that the evaluation provides them with nothing to enable them to perform their duties more effectively. Frequent feedback can overcome this understandable resistance.

Multifaceted evaluation designs also require the use of a variety of quantitative and qualitative data collection methods. Evaluations might include surveys and interviews, focus groups, archival research, and observations. Evaluator-practitioner teams should consider all these techniques in light of their appropriateness to the particular intervention and their strengths and weaknesses.

The final topic in this guide is data analysis. The differences among several types of descriptive and relational analyses on quantitative data were discussed. Without delving into the specifics of how to perform complex statistical analysis, the conceptual underpinnings of these types of analysis were presented in the belief that prevention administrators should be able not only to understand the results of data analysis, but also to ask better questions about the plans for data collection and analysis before the evaluation is actually implemented. The steps in qualitative data analysis were also discussed to guide the transformation of volumes of qualitative data into accessible and useful information.

The potential benefits of employing the recommendations discussed in this guide are many. Evaluation will become an ongoing, dynamic, collaborative process. Evaluation expectations will be clear and appropriate. Information will steer program development. All prevention activities will be monitored and evaluated at some level, but the most costly outcome evaluation activities will be reserved for those activities that are truly ready for outcome evaluation. Using this structure for collaborative evaluation, prevention programs can expect to strengthen their interventions and to amass solid evidence of their effectiveness.

References

- Bangert-Drowns, R. L. (1988). The effects of school-based substance abuse education—A meta-analysis. *Drug Education* 18(3), 243–264.
- Bureau of Justice Assistance. (1993). *Problem-oriented drug enforcement: A community-based approach to effective policing*. Washington, DC: Author.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Clayton, R. R., Cattarello, A., Day, L. E., & Walden, K. P. (1991). Persuasive communication and drug prevention: An evaluation of the DARE program. In L. Donohew, H. E. Sypher, & W. Bukoski (Eds.), *Persuasive communication and drug abuse prevention* (pp. 295–313). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Davis, R. C., & Lurigio, A. J. (1996). *Fighting back: Neighborhood anti-drug strategies*. Thousand Oaks, CA: Sage.
- Florin, P., Mitchell, R., & Stevenson, J. (1993). Identifying training and technical assistance needs in community coalitions: A developmental approach. *Health Education Research* 8(3), 417–432.
- French, W. L., & Bell, C. H., Jr. (1984). *Organization development: Behavioral science interventions for organization improvement* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Goodman, R. M., & Wandersman, A. (1994). FORECAST: A formative approach to evaluating community coalitions and community-based initiatives. *Journal of Community Psychology* (Monograph series, CSAP special issue), 6–25.
- Gottfredson, G. D. (1984). A theory-ridden approach to program evaluation: A method for stimulating researcher-implementer collaboration. *American Psychologist* 39(10), 1101–1112.
- Green, L. (1996). *Policing places with drug problems* (Drug, health, and social policy series, Vol. 2). Thousand Oaks, CA: Sage.
- Holder, H., Saltz, R. F., Grube, J. W., Treno, A. J., Reynolds, R. I., Voas, R. B., & Gruenewald, P. J. (1997). Summing up: Lessons from a comprehensive community prevention trial. *Addiction* 92(Suppl. 2), S293–S301.
- Kibel, B. (1994). *Open-systems planning-for-action: A practical approach to planning, implementing, and evaluating community change processes*. Unpublished meeting handout available from author.
- Moskowitz, J. M. (1989). The primary prevention of alcohol problems: A critical review of the research literature. *Journal of Studies on Alcohol* 50(1), 54–88.
- National Institute on Drug Abuse. (1988). *Working with evaluators: A guide for drug abuse prevention program managers*, J. French, C. Fischer, & S. Costa, Jr. (Eds.). (DHHS Publication No. ADM 88-1233). Rockville, MD: Author.
- National Institute on Drug Abuse. (1991). *Drug abuse prevention intervention research: Methodological issues*, C. G. Leukefeld & W. J. Bukoski (Eds.). (NIDA Research Monograph 107, DHHS Publication No. ADM 91-1761). Rockville, MD: Author.
- Pentz, M., Dwyer, J. H., MacKinnon, D. P., Flay, B. R., Hansen, W. B., Wang, E. Y., & Johnson, C. A. (1989). A multicomunity trial for primary prevention of adolescent drug abuse: Effects on drug use prevalence. *Journal of the American Medical Association* 261(22), 3259–3266.
- Perry, C. L., Williams, C. L., Veblen-Mortensen, S., Toomey, T. L., Anstine, P., McGovern, P., Finnegan, J. P., Wagenaar, A. C., & Wolfson, M. (1996). Project Northland: Outcomes of a community-wide alcohol use prevention program during early adolescence. *American Journal of Public Health* 86(7), 956–965.
- Ringwalt, C., Ennett, S. T., & Holt, K. D. (1991). An outcome evaluation of project DARE (Drug Abuse Resistance Education). *Health Education Research* 6(3), 327–337.

- Roehl, J. A., Wong, H., Andrews, C., Huitt, R., & Capowich, G. E. (1995). *A national assessment of community-based anti-drug initiatives: Final report submitted to the National Institute of Justice*. Pacific Grove, CA: Institute for Social Analysis.
- Rosenbaum D. P., Flewelling, R. L., Bailey, S. L., Ringwalt, C. L., & Wilkinson, D. L. (1994). Cops in the classroom: A longitudinal evaluation of drug abuse resistance education (DARE). *Journal of Research in Crime and Delinquency* 31(1), 3–31.
- Rundall, T. G., & Bruvold, W. H. (1988). A meta-analysis of school-based smoking and alcohol use prevention programs. *Health Education Quarterly* 15(3), 317–334.
- Schroeder, D. S., Laflin, M. T., & Weis, D. L. (1993). Is there a relationship between self-esteem and drug use? Methodological and statistical limitations of the research. *Journal of Drug Issues* 23(4), 645–665.
- Sechrest, L., & Figueredo, A. J. (1993). Program evaluation. In L. W. Porter & M. R. Rosenzweig (Eds.), *Annual Review of Psychology*. Vol. 44 (pp. 645–674). Palo Alto, CA: Annual Reviews, Inc.
- Tobler, N. S. (1992). Drug prevention programs can work: Research findings. *Journal of Addictive Diseases* 11(3), 1–28.
- Wagenaar, A. C., Murray, D. M., Wolfson, M., Forster, J. L., & Finnegan, J. R. (1994). Communities mobilizing for change on alcohol: Design of a randomized community trial. *Journal of Community Psychology* (Monograph series, CSAP special issue), 79–101.
- Selected Related SAMHSA/CSAP Publications**
Advanced Methodological Issues in Culturally Competent Evaluation for Substance Abuse Prevention. (1996). CSAP Cultural Competence Series 6. 266 pp. DHHS Pub. No. (SMA)96-3110.
- Experiences With Community Action Projects: New Research in the Prevention of Alcohol and Other Drug Problems*. (1993). CSAP Prevention Monograph 14. 301 pp. DHHS Pub. No. (ADM)93-1976.
- Handbook for Evaluating Drug and Alcohol Prevention Programs*. (1987). 113 pp. DHHS Pub. No. (ADM)87-1512.
- Journal of Community Psychology*. (1994). Monograph Series, CSAP Special Issue. See: Cook, R., Roehl, J., Oros, C., and Trudeau, J. Conceptual and methodological issues in the evaluation of community-based substance abuse prevention coalitions: Lessons learned from the National Evaluation of the Community Partnership Program. pp. 155–169.
- Prevention Plus III: Assessing Alcohol and Other Drug Prevention Programs at the School and Community Level*. (1991). 461 pp. DHHS Pub. No. (ADM)91-1817.
- Research, Action, and the Community: Experiences in the Prevention of Alcohol and Other Drug Problems*. (1990). OSAP Prevention Monograph 4. 326 pp. DHHS Pub. No. (ADM)89-1651.

Appendix

Resources on Evaluation

- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Gottfredson, D. C., Hawkins, J. D., & Netherhood, B. (1990). *Handbook for evaluating drug and alcohol prevention programs: A guide prepared for use in the Maryland evaluation of drug prevention programs for high-risk youths*. College Park, MD: University of Maryland, Institute of Criminal Justice and Criminology.
- Hays, W. L. (1988). *Statistics* (4th ed.). Fort Worth, TX: Holt, Rinehart, and Winston.
- Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart, and Winston.
- Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research*. Qualitative Research Methods Series 1. Beverly Hills, CA: Sage.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- National Institute on Drug Abuse. (1981). *Handbook for prevention evaluation: Prevention evaluation guidelines*. J. F. French & N. J. Kaufman (Eds.). DHHS Pub. No. (ADM)81-1145. Rockville, MD: Author.
- National Institute on Drug Abuse. (1994a). *Advances in data analysis for prevention intervention research* (NIDA Research Monograph 142). L. M. Collins & L. A. Seitz (Eds.). NIH Pub. No. 94-3599. Rockville, MD: Author.
- National Institute on Drug Abuse. (1994b). *Scientific methods for prevention intervention research* (NIDA Research Monograph 139). A. Cazares & L. A. Beatty (Eds.). NIH Pub. No. 94-3631. Rockville, MD: Author.
- National Institute on Drug Abuse. (1995). *Qualitative methods in drug abuse and HIV research* (NIDA Research Monograph 157). E. Y. Lambert, R. S. Ashery, & R. H. Needle (Eds.). NIH Pub. No. 95-4025. Rockville, MD: Author.
- Pagano, R. R. (1986). *Understanding statistics in the behavioral sciences* (2nd ed.). St. Paul, MN: West.
- Patton, M. Q. (1996). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Rossi, P., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th ed.). Newbury Park, CA: Sage.
- Wolcott, H. F. (1990). *Writing up qualitative research* (Qualitative Research Methods Series 20). Newbury Park, CA: Sage.

Glossary

Archival research: Research method that involves the use of data extracted from existing written or computer records.

Attribution of effect: Concept that refers to the ability to conclude that an intervention *caused* an outcome.

Comparison group: In quasi-experimental evaluation design, a group of evaluation participants that is not exposed to the intervention. This term usually implies that participants were *not* randomly assigned, but were similar to the intervention group members in many respects.

Confidence interval: An estimated range of values derived from sample statistics with a given high probability of covering the true population value.

Contingency table: A form of relational analysis that classifies observations by their values on two or more variables. Each cell in the table represents a unique combination of values across the variables, and each observation qualifies for one and only one cell in the table.

Control group: In experimental evaluation design, a group of participants that is essentially similar to the intervention group but is not exposed to the intervention. Participants are designated to be part of either a control or intervention group through random assignment.

Convenience sample: A nonrandom study sample selected on the basis of convenience and accessibility.

Correlational analysis: A form of relational analysis that assesses the strength and direction of association between variables.

Data: Information collected according to a methodology and through specific research methods and instruments.

Data analysis: The process of examining systematically collected information.

Descriptive analysis: Data analysis that results in information that characterizes the sample, such as measures of central tendency (e.g., mean, median, mode) and measures of variability (e.g., range, standard deviation, variance). In addition to describing the sample, such data may be used as input for relational analysis.

Experimental design: A methodology for examining intervention outcomes that involves the random assignment of subjects to intervention and control conditions with a controlled manipulation delivered to subjects in the intervention group. The design enables the evaluator to conclude that the outcomes were *caused* by the intervention. See also **quasi-experimental design**.

Face-to-face questionnaire: See **structured interview**.

Focus group: Qualitative research method that involves structured discussion among individuals with shared characteristics.

Individual-oriented intervention: An intervention that attempts to change the behavior of individuals by enhancing the knowledge, attitudes, skills, and beliefs of individuals.

Informed consent: The written permission obtained from research participants (or their parents if participants are minors) giving their consent to participate in an evaluation after having been informed of the nature of the research.

Institutional review board: A group of researchers and others appointed by an institution to assess proposed data collection regarding potential harm that might be caused to study participants.

Instrument: Device that assists evaluators in collecting data in an organized fashion, such as a standardized survey or interview protocol.

Intermediate outcome: Intervention outcome, such as changes in knowledge, attitudes, or beliefs, that occurs before, and is necessary for changes in, substance use and substance-related problems. See also **long-term outcome**.

Internal validity: Concept that refers to the ability to make inferences about whether the relationship between variables is causal in nature and, if it is, the direction of causality.

Intervention: A manipulation applied to a population in order to change behavior. In substance abuse prevention, intervention at the individual or policy level may be used to prevent or lower the rate of substance use or related problems. See also **policy intervention**; **individual-oriented intervention**.

Intervention group: In experimental and quasi-experimental evaluation designs, the group of participants that is exposed to the intervention. See also **control group** and **comparison group**.

Long-term outcome: Intervention outcome corresponding to the prevalence of substance use and substance-related problems.

Maturation effects: Changes in outcomes that are attributable to participants' growing older, wiser, stronger, more experienced, and the like, solely through the passage of time.

Mean: A measure of central tendency computed by summing over all the values of a variable and dividing by the number of cases (on average).

Means comparison: A form of relational analysis that involves comparing the average values of two or more groups to see if they differ more than would be expected by chance.

Measures of central tendency: Indices that describe the "typical" or "average" value—for example, the arithmetic average and the median.

Median: A measure of central tendency referring to the point exactly midway between the top and bottom halves of a distribution of values.

Methodology: Procedure for collecting data.

Mode: A measure of central tendency referring to the value most often given by respondents.

Nonrepresentative sample: A segment of a larger body or population that does not mirror in composition characteristics of the larger body or population. See also **representative sample**.

Observation: Data collection method involving unobtrusive examination of behavior and/or occurrences, often in a natural setting, and characterized by no interaction between participants and observers.

Outcome evaluation: Evaluation that focuses research questions on assessing intervention effects on intended outcomes. See also **process evaluation**.

Participant-observation: Qualitative research method that requires simultaneous participation in and examination of activity in a natural setting. The identity of the evaluator as an evaluator is usually made known to others in the setting.

Policy intervention: An intervention that attempts to change the behavior of individuals by changing the economic or regulatory environment around substance use.

Process evaluation: Evaluation that focuses research questions on the nature of intervention implementation or its structure and operations. See also **outcome evaluation**.

Qualitative data: Generally, contextual information in evaluation studies, usually describing participants and interventions. The strength of qualitative data, which is often presented as text, is its ability to illuminate evaluation findings derived from quantitative methods. See also **quantitative data**.

Quantitative data: Measures that in evaluation studies capture changes in targeted outcomes (e.g., substance use) and intervening variables (e.g., attitudes toward use). The strength of quantitative data is its use in testing hypotheses and determining the strength and direction of effects.

Quasi-experimental design: A plan for examining intervention outcomes that involves an intervention group and may involve a comparison group and/or preintervention and/or postintervention tests. This design does not involve random assignment of participants to conditions.

Questionnaire: Research instrument that consists of written questions, each with a limited set of possible responses. See also **self-administered questionnaire**; **structured interview**; **telephone survey**.

Random assignment: The process through which members of a pool of eligible evaluation participants are assigned to either the intervention group or a control group on a random basis, such as through the use of a table of random numbers.

Relational analysis: Data analysis that reveals the relationship between variables considered important for the evaluation—for example, correlational analysis and regression analysis.

Representative sample: A segment of a larger body or population that mirrors in composition the characteristics of the larger body or population. See also **nonrepresentative sample**.

Sample: A segment of a larger body or population.

Sample attrition: Unplanned reduction in the size of the study sample because of participants' dropping out of the evaluation—for example, because of relocation.

Self-administered questionnaire: A questionnaire that is completed by the respondent without any assistance or clarification from the evaluator.

Self-selection: An occurrence in which individuals themselves choose to participate in a program or become a member of a sample without the control of the evaluator.

Semistructured interview: Qualitative data collection method that involves an interviewer and specific questions with unlimited response options. See also **structured interview**; **unstructured interview**.

Simple random sample: In experimental research designs, a sample derived from indiscriminate selection from a pool of eligible participants, such that each member of the population has an equal chance of being selected for the sample. See also **stratified random sample**.

Snowball sample: A nonrandom sample that is composed according to the referrals of initial sample members such that sample members not only share certain common characteristics, but are likely to be familiar with one another. Also referred to as chain sample.

Standard deviation: A unit of measure of variability or dispersion characterizing the tendency for observations to depart from central tendency. Standard deviation and its square, variance, reflect how accurately the central tendency measures (such as the mean) would describe a randomly selected observation.

Statistical significance: A term referring to the strength of a particular relationship between variables. A relationship is said to be statistically significant when it occurs so frequently in the data that the relationship's existence is probably not attributable to chance.

Stratified random sample: In experimental research designs, a sample group derived from indiscriminate selection from different subsegments of a pool of eligible participants (e.g., men, women). See also **simple random sample**.

Structured interview: Quantitative data collection method that involves an interviewer, specific questions, and limited sets of possible responses to each question. Sometimes referred to as face-to-face questionnaire. See also **semistructured interview**; **unstructured interview**.

Telephone survey: A structured interview conducted over the telephone.

Time-series analysis: A form of data analysis that involves examination of data derived from repeated assessments across time.

Unstructured interview: Qualitative data collection method that involves an interviewer and given questions. Not all given questions may be asked, however, and additional substantive questions (not necessarily questions for clarification purposes) may be posed by the interviewer.

Variable: Factor or characteristic of the intervention, participant, and/or context that may influence or be related to the possibility of achieving intermediate and long-term outcomes.

NIH LIBRARY



3 1496 01031 0277